

(19) World Intellectual Property  
Organization  
International Bureau



(43) International Publication Date  
4 August 2005 (04.08.2005)

PCT

(10) International Publication Number  
**WO 2005/069859 A2**

(51) International Patent Classification: Not classified

(21) International Application Number:  
PCT/US2005/001218

(22) International Filing Date: 13 January 2005 (13.01.2005)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
60/536,337 13 January 2004 (13.01.2004) US

(71) Applicant (*for all designated States except US*): **UNIVERSITY OF UTAH RESEARCH FOUNDATION** [US/US];  
615 Arapeen Drive, Suite 310, Salt Lake City, UT 84108 (US).

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): **THUMMEL, Carl, S.** [US/US]; 2352 S. Lakeline Drive, Salt Lake City, UT 84109 (US). **KING-JONES, Kirst** [GB/US]; 1416 Downington Avenue, Salt Lake City, UT 84105 (US). **HORNER, Michael** [US/US]; 1619 E. Wilson Avenue, Salt Lake City, UT 84105 (US). **LAM, Geanette** [US/US]; 4984 S. Kalani Drive, Holliday, UT 84117 (US).

(74) Agents: **CLEVELAND, Janell, T.** et al.; Needle & Rosenberg, P.C., Suite 1000, 999 Peachtree Street, Atlanta, GA 30309-3915 (US).

(81) Designated States (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GII, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

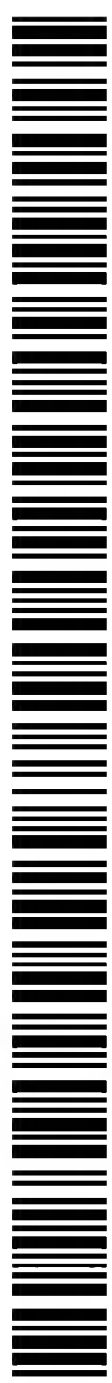
**Published:**

— *without international search report and to be republished upon receipt of that report*

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

(54) Title: COMPOSITIONS AND METHODS FOR MODULATING DHR96

(57) Abstract: Disclosed are compositions and methods for modulating DHR96 activity and identifying molecules that modulate DHR96 activity.



**WO 2005/069859 A2**

**COMPOSITIONS AND METHODS FOR MODULATING DHR96****I. BACKGROUND**

1. The control of insects with toxins (pesticides) is one of the largest industries in the world. Insects have evolved many methods to deal with pesticides, most of which act through a xenobiotic detoxification pathway. The regulation of the xenobiotic pathway represents an attractive target for pesticides. Disclosed herein, DHR96, a *Drosophila* gene is shown to regulate the xenobiotic pathway, and inhibition of the DHR96 gene expression or activity decreases the ability of *Drosophila* to adapt to toxins, including pesticides, such as DDT.

**II. SUMMARY**

2. Disclosed are methods and compositions related to compositions and methods for regulating DHR96 and increasing the effect of existing any toxins to control insects are disclosed.

**III. BRIEF DESCRIPTION OF THE DRAWINGS**

3. The accompanying drawings, which are incorporated in and constitute a part of this specification, illustrate several embodiments and together with the description illustrate the disclosed compositions and methods.

4. Figure 1 shows DHR96 is closely related to the PXR/CAR/VDR subfamily of xenobiotic receptors. An alignment using the programs PHYLIP and CLUSTALW is depicted of the DHR96, DAF-12, PXR, CAR, and NHR-8 nuclear receptors, showing the percent identical amino acids within either the DNA binding domain or ligand binding domain.

5. Figure 2 shows DHR96 is expressed in organs involved in nutrient absorption, metabolism, and excretion. Organs were dissected from wandering third instar larvae, fixed in 25% formaldehyde and stained with affinity-purified antibodies to detect DHR96 protein. In wild type larvae, nuclear DHR96 protein is detected in the fat body, in salivary glands and regions of the digestive tract including the gastric caeca and the Malpighian tubules. Only background staining is detected in other tissues, including the imaginal discs and brain. No expression was detectable in fat bodies dissected from *DHR96*<sup>E25</sup> mutant larvae, demonstrating the specificity of the antibody stains.

6. Figure 3 shows a strategy for targeted mutagenesis of the DHR96 locus.  $\Delta 1$  depicts the start methionine deletion and  $\Delta 2$  depicts the deletion of the fourth exon/intron of *DHR96*. A transgene containing the targeting construct and the GFP marker was circularized by FLP recombinase and subsequently cut with I-SceI. Homologous pairing between the targeting

construct and the endogenous *DHR96* locus results in the generation of a tandem duplication by 'ends-in' recombination. To generate a single copy insertion, the tandem duplication was reduced by means of homologous recombination by inducing a DNA double stranded break with I-CreI.

7. Figure 4 shows DHR96 mutants are more sensitive than wild type flies to the pesticide DDT. A time course is shown. 20 wild type or *DHR96*<sup>E25</sup> mutant flies were treated with a high concentration of DDT (100 ng/μl) and assayed for survival every hour up to 10 hours. Each assay (A+B) was done in triplicate to determine the standard deviation as shown by the error bars.

8. Figure 5 shows an alignment of Drosophila nuclear hormone receptor DNA-binding domains. An alignment of the DNA-binding domains of known Drosophila nuclear hormone receptor superfamily members reveals two regions of conserved amino acids flanking a central unique region. The conserved amino acids were used to design PCR primers for amplifying fragments of Drosophila receptors: F3, F4, F5, R4, R5, R6 and R8. The unique region was used to design gene-specific oligonucleotide probes to eliminate previously identified family members from further study.

9. Figure 6 shows alignments of DNA-binding domain sequences. The DNA-binding domain sequence of each gene was used to search the PIR/Swiss Prot/GenBank databases. An alignment of each sequence with representative matches from the databases is presented. Shaded boxes indicate identity with the new protein sequence, and the percent identity is shown to the right of each sequence.

10. Figure 7 shows temporal profiles of DHR38, DHR78, and DHR96 transcription during the onset of metamorphosis. Northern blots containing RNA samples isolated from staged third instar larvae and prepupae collected at 2 hr intervals were probed to detect DHR38, DHR78, and DHR96 mRNAs. These blots have been used previously for detailed studies of 20E-regulated gene transcription ((Andres, A. J., Fletcher, J. C., Karim, F. D. & Thummel, C. S. (1993). Dev. Biol. 160, 388-404) One set of blots was sequentially stripped and hybridized with probes from each gene, in order to allow direct comparison of transcription patterns. The blots were also hybridized to detect rp49 mRNA, as a control for equal loading (data not shown)). Developmental times are shown at the top as hours after egg laying for third instar larval development, and as hours after puparium formation for prepupal and pupal development. Landmark 20E-triggered developmental transitions are shown at the top.

11. Figure 8 shows a time course of DHR38, DHR78, and DHR96 transcription in cultured larval organs treated with 20E. Mass-isolated late third instar larval organs were treated with  $5 \times 10^{-7}$  M 20E for the times shown, as described (Thummel, C. S., Burtis, K. C. & Hogness, D. S. (1990). Cell 61, 101-111) Equal amounts of total RNA isolated from each time point were fractionated by formaldehyde agarose gel electrophoresis, transferred to a nylon membrane, and hybridized with probes to detect DHR38, DHR78, DHR96 and rp49 mRNA. One northern blot was sequentially stripped and hybridized with a probe from each gene, in order to allow direct comparison of transcription patterns. Detection of DHR38 transcripts required the use of an antisense RNA probe.

12. Figure 9 shows the DNA-binding specificities of DHR38, DHR78, and DHR96 protein. Each protein was overproduced in *E. coli*, purified, and tested for its ability to bind to eight oligonucleotides using electrophoretic mobility shift assays. The names of each oligonucleotide are shown at the top. In all cases, binding could be competed by the addition of an excess of the appropriate unlabelled oligonucleotide. Figure 10 shows that no DHR96 protein was detectable in DHR96 mutants. Total protein was isolated from wild type control flies (w<sup>1118</sup>) DHR96E25 mutants, DHR9616A mutants, or 1/50 the amount of protein from heat-induced hs-DHR96 transformants that overexpress DHR96 protein were analyzed on a Western blot using DHR96 antibodies. The mutants shown in the center two lanes had no detectable DHR96 protein.

13. Figure 10 shows DHR96E25 mutants are sensitive to phenobarbital and tebufenozide. Control Canton S adult flies (CanS), original DHR96E25 mutants (DHR96E25), and the outcrossed DHR96E25 mutant (outcross 1) were exposed to either DDT (Fig. 11A) or phenobarbital (Fig. 11B) for 23 hours and then scored for viability or motility, respectively. A dose response curve is shown. Twenty wild type or *DHR96*<sup>E25</sup> mutant flies were exposed to eight DDT concentrations, from 0.78 to 100 ng/ $\mu$ l, and then scored for survival 10 hours later. A similar test was conducted for sensitivity to tebufenozide (Fig. 11C) using larvae raised on food supplemented with the drug. In parallel experiments, the original DHR9616A stock showed responses similar to the original DHR96E25 mutant.

14. Figure 11 shows that *DHR96* regulates members of all four classes of insect detoxification genes. The top genes that are down-regulated upon ectopic DHR96 overexpression are listed. Total RNA was extracted and purified to allow probe generation. Affymetrix microarray chips were hybridized with the probes and scanned. Raw data was analyzed with dCHIP, and filtering was performed in MS ACCESS. The expression levels in

control (WWPHS) and *hs-DHR96* (96WPHS) animals are shown, along with the fold change in gene expression. Members of gene families known to be involved in detoxification in insects are also shown.

15. Figure 12 shows a schematic representation of the GAL4-LBD activation assay. A gene fusion of the GAL4 DNA binding domain (DBD) and DHR96 ligand binding domain (LBD) is expressed upon heat-induction of the *hsp70* promoter. The resultant fusion protein can bind to GAL4 response elements (UAS) on a separate transgenic construct, but will only activate *lacZ* transcription in the presence of an appropriate ligand and/or co-factors (a ligand is shown).  $\beta$ -galactosidase expression is detected as the substrate from an Xgal staining reaction.

16. Figure 13 shows GAL4-DHR96 is activated by tebufenozide. Third instar larvae were heat-treated to induce GAL4-DHR96 expression, dissected, and organs were cultured in the presence of  $1 \times 10^{-5}$  M tebufenozide. UAS-lacZ reporter gene expression was detected by Xgal staining. Control animals were either from a non-transgenic control line or GAL4-DHR96 transgenic animals that were not treated with tebufenozide.

#### IV. DETAILED DESCRIPTION

17. Before the present compounds, compositions, articles, devices, and/or methods are disclosed and described, it is to be understood that they are not limited to specific synthetic methods or specific recombinant biotechnology methods unless otherwise specified, or to particular reagents unless otherwise specified, as such may, of course, vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only and is not intended to be limiting.

##### A. Definitions

18. As used in the specification and the appended claims, the singular forms "a," "an" and "the" include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to "a pharmaceutical carrier" includes mixtures of two or more such carriers, and the like.

19. Ranges can be expressed herein as from "about" one particular value, and/or to "about" another particular value. When such a range is expressed, another embodiment includes from the one particular value and/or to the other particular value. Similarly, when values are expressed as approximations, by use of the antecedent "about," it will be understood that the particular value forms another embodiment. It will be further understood that the endpoints of each of the ranges are significant both in relation to the other endpoint, and independently of the

other endpoint. It is also understood that there are a number of values disclosed herein, and that each value is also herein disclosed as “about” that particular value in addition to the value itself. For example, if the value “10” is disclosed, then “about 10” is also disclosed. It is also understood that when a value is disclosed that “less than or equal to” the value, “greater than or equal to the value” and possible ranges between values are also disclosed, as appropriately understood by the skilled artisan. For example, if the value “10” is disclosed the “less than or equal to 10” as well as “greater than or equal to 10” is also disclosed. It is also understood that the throughout the application, data is provided in a number of different formats, and that this data, represents endpoints and starting points, and ranges for any combination of the data points.

For example, if a particular data point “10” and a particular data point 15 are disclosed, it is understood that greater than, greater than or equal to, less than, less than or equal to, and equal to 10 and 15 are considered disclosed as well as between 10 and 15.

20. References in the specification and concluding claims to parts by weight, of a particular element or component in a composition or article, denotes the weight relationship between the element or component and any other elements or components in the composition or article for which a part by weight is expressed. Thus, in a compound containing 2 parts by weight of component X and 5 parts by weight component Y, X and Y are present at a weight ratio of 2:5, and are present in such ratio regardless of whether additional components are contained in the compound.

21. A weight percent of a component, unless specifically stated to the contrary, is based on the total weight of the formulation or composition in which the component is included.

22. In this specification and in the claims which follow, reference will be made to a number of terms which shall be defined to have the following meanings:

23. “Optional” or “optionally” means that the subsequently described event or circumstance may or may not occur, and that the description includes instances where said event or circumstance occurs and instances where it does not.

24. “Primers” are a subset of probes which are capable of supporting some type of enzymatic manipulation and which can hybridize with a target nucleic acid such that the enzymatic manipulation can occur. A primer can be made from any combination of nucleotides or nucleotide derivatives or analogs available in the art which do not interfere with the enzymatic manipulation.

25. “Probes” are molecules capable of interacting with a target nucleic acid, typically in a sequence specific manner, for example through hybridization. The hybridization of nucleic acids

is well understood in the art and discussed herein. Typically a probe can be made from any combination of nucleotides or nucleotide derivatives or analogs available in the art.

26. Throughout this application, various publications are referenced. The disclosures of these publications in their entireties are hereby incorporated by reference into this application in order to more fully describe the state of the art to which this pertains. The references disclosed are also individually and specifically incorporated by reference herein for the material contained in them that is discussed in the sentence in which the reference is relied upon.

## **B. Compositions and methods**

27. Four lines of evidence show that DHR96 plays a central role in coordinating insect xenobiotic responses. First, this gene is a member of the nuclear receptor subclass that includes the PXR, SXR, VDR, and NHR-8 xenobiotic receptors. Second, DHR96 protein is expressed specifically in tissues that are involved in absorption, metabolism, and excretion of toxic compounds. Third, a *DHR96* mutant is sensitive to phenobarbital and tebufenozide. Finally, members of all four classes of known insect detoxification genes can be regulated by ectopic DHR96 expression.

28. Higher organisms neutralize environmental toxins or xenobiotics through enzymes that include cytochrome p450 monooxygenases, glutathione transferases, carboxylesterases, and UDP-glucuronosyl transferases. In mammals, some of these detoxification enzymes are directly regulated by the nuclear receptors PXR and CAR, which in turn are activated by a broad spectrum of xenobiotics including prescription drugs, plant toxins and other contaminants. In contrast, there is little understanding of how similar xenobiotic responses might be controlled in insects. Herein it is shown that mutants in the DHR96 nuclear receptor of *Drosophila* are viable and fertile under standard laboratory conditions, as are flies that widely express double stranded *DHR96* RNA (RNAi) from a transgene. However, when exposed to a pesticide like DDT, mutant animals are less resistant to the insecticide challenge, dying more rapidly and at lower concentrations than control animals. Unlike many other nuclear receptors, widespread ectopic expression of DHR96 has no effect on the viability of larvae or flies, suggesting that activation of DHR96 is ligand-dependent.

29. Disclosed herein, DHR96 is expressed in tissues that have been associated with the detoxification process, including the gastric caeca, the major site of absorption in Diptera, and the fat body, the insect equivalent of the liver. Microarray studies disclosed herein show that overexpression of *DHR96* results in the downregulation of members of all four classes of the

detoxification machinery, supporting the proposal that DHR96 functions as a xenobiotic regulator in *Drosophila*. These findings demonstrate how detoxification enzymes are activated in insects upon challenge with an insecticide. Given that this receptor has been highly conserved in the distant insect species, *Anopheles gambiae*, it is likely that it exerts a similar function in all insects. Also disclosed are methods for the identification of specific compounds or peptides that affect DHR96 activity and can act as effective synergists that, for example, enhance the lethality of pesticides for insect control.

30. Disclosed are mutants of the DHR96 gene which have reduced DHR96 activity in the xenobiotic pathway. These mutants can be used in a variety of methods for isolating new molecules that inhibit the xenobiotic pathway, by for example, being used as controls in methods that are testing the xenobiotic activity of a particular compound. The mutants can also be used as stock for production of other mutant flies. The mutants can also be used as seed genetic backgrounds to change a given population of flies to insecticide sensitive flies, by introducing the mutant backgrounds into the populations, through fly breeding.

31. Also disclosed are compositions which are capable of inhibiting DHR96 protein function or gene function, and which in turn inhibit the xenobiotic effect of the DHR96 protein. For example, disclosed are iRNA molecules which inhibit the function of DHR96 and inhibit the xenobiotic effect of DHR96.

32. Also disclosed are methods of inhibiting insect growth by administering an inhibitor of DHR96 to an insect, such as a fly.

33. Also disclosed are methods of identifying molecules that inhibit DHR96, and inhibit the xenobiotic activity in an insect, such as a fly, comprising for example, testing compounds for inhibition activity of DHR96 and/or inhibition of xenobiotic activity and, then for example, comparing the activity of these molecules to the disclosed inhibitors of DHR96, such as the mutants or the disclosed iRNA molecules.

### 1. The xenobiotic response

34. Virtually every organism faces a fundamental challenge when exposed to potentially harmful environmental substances called xenobiotics, which may include pharmaceuticals, plant toxins, pollutants, pesticides, hormones and fatty acids. Exposure to xenobiotics can occur either directly by physical contact, inhalation, or ingestion of nutrients or indirectly when an organism generates toxic metabolites from less harmful precursors. The mechanisms by which toxic compounds are removed and/or neutralized fall into two broad categories. Usually as a result of extreme selective pressures, organisms may develop adaptive processes that are highly specific

to a particular substance, as can be observed in many insect species that become resistant to pesticides (Wilson, T. G. (2001). *Annu Rev Entomol* 46, 545-571) or that have evolved the ability to utilize hazardous plant species as a food source (Danielson, P. B. et al. (1997). *Proc Natl Acad Sci U S A* 94, 10797-10802; Fogleman, J. C. (2000). *Chem Biol Interact* 125, 93-105.). In contrast to this highly specific response, all metazoan species appear to have a general machinery that allows the efficient detoxification of a vast range of chemicals. The general detoxification mechanisms display a surprising flexibility, which is mainly achieved by two factors. First, at least three enzyme classes comprising more than 160 proteins in the mosquito and the fruit fly are responsible for metabolizing lipophilic toxins into less harmful substances (Ranson, H., et al. (2002). *Science* 298, 179-181). Second, some enzymes appear to have an immense range of substrate specificity. For instance, Cyp3A4, a member of the cytochrome p450 monooxygenase family, is capable of neutralizing an estimated 50% of all existing prescription drugs (Maurel, P. (1996). (Boca Raton, CRC Press), pp. 241-270). Cytochrome p450 enzymes are often referred to as phase I enzymes, because they catalyze the first step in the detoxification process by adding oxygen groups to lipophilic chemicals, thus resulting in more water-soluble compounds, which in turn facilitates efficient excretion. Other enzyme families like glutathione transferases, carboxylesterases and UDP-glucuronosyl transferases are classified as phase II enzymes, as their role is to catalyze subsequent detoxification steps.

35. In insects, pesticide resistance is most often the result of mutations that affect the general detoxification pathway. For example, the overexpression of a single gene, *Cyp6g1*, a member of the cytochrome p450 family, is sufficient to confer DDT resistance in *Drosophila melanogaster* (Daborn, P. B. et al. (2002), *Science* 297, 2253-2256). The same study demonstrated that *Cyp6g1* is hypertranscribed in over 20 DDT-resistant *Drosophila* strains of worldwide origin, but further analysis suggested that this finding could be traced back to a single event, since all alleles harbor the same *Accord* transposon in their 5' regulatory region.

36. In the past decade considerable progress in the field has revealed the mechanisms that allows an organism to sense a wide range of toxic substances and to understand how xenobiotic sensing translates into the induction of highly specific sets of detoxifying enzymes. It quickly became apparent that certain members of the so-called nuclear receptor superfamily are the central players in this process. Nuclear receptors are ligand-activated transcription factors that play important roles in diverse physiological processes such as cell growth and differentiation, embryonic development, and cholesterol metabolism (Francis, G. A. et al. (2003) *Annu Rev Physiol* 65, 261-311; Mangelsdorf, D. J., et al. (1995). *Cell* 83, 835-839; Tontonoz, P., and

Mangelsdorf, D. J. (2003). Mol Endocrinol 17, 985-993) Of the 48 nuclear receptors encoded by the human genome ~26 have identified ligands (Kliwer, S. A. (2003) J Nutr 133, 2444S-2447S), but only three have been associated with xenobiotic activity, namely PXR, CAR and VDR (Maglich, J. M., et al. (2002) Mol Pharmacol 62, 638-646; Makishima, M., et al.

(2002). Science 296, 1313-1316). These three closely related receptors are not only able to sense and bind lipophilic xenobiotic substances directly, but once activated by such a ligand, they can regulate the expression of enzymes that will neutralize the very compound that had activated these nuclear receptors in the first place, thus creating feedback loop. Disclosed is an analogous mechanism that exists in the fruit fly, *Drosophila melanogaster*. The disclosed mechanism involves an insect nuclear receptor, the Drosophila DHR96 nuclear receptor.

### (1) Nuclear receptors

37. Members of the nuclear receptor superfamily have been one of the most productive targets for drug development by the pharmaceutical industry. Efforts along these lines have resulted in drugs that have had a major impact on human health, including cancer treatments, fertility control, and cholesterol reduction. Nuclear receptors are ligand-activated transcription factors, but can have many regulatory functions aside from this ligand activated function. Nuclear receptors have been organized in a phylogeny-based nomenclature (Nuclear Receptors Nomenclature Committee, (1999) Cell 97, 1-3.) of the form NR<sub>xyz</sub>, where x is the sub-family, y is the group and z the gene. For a review see, Robinson-Rechavi, M., et al., Journal of Cell Science, Cell Science at a Glance, 116(4):585-586 and poster insert, (2003), which is herein incorporated by reference at least for material related to nuclear receptors).

38. Nuclear receptors lend themselves to drug intervention because their activity can be modulated by small lipophilic compounds that can be easily delivered to animals in a stable format. Compounds can be developed that either constitutively activate their cognate receptor, called agonists, or constitutively inactivate the receptor, called antagonists. The use of these compounds in animals provides a means of tightly regulating nuclear receptor activity *in vivo*, with resultant effects on growth and development.

39. Surprisingly, no similar effort has been made by the agricultural industry to target insect nuclear receptors as a means of pest control. This is largely because the mechanism of action of most insect nuclear receptors has remained undefined. Disclosed herein it was shown that an insect nuclear receptor, encoded by *DHR96*, is required for resistance to toxic compounds in *Drosophila*. Also disclosed are molecules that inhibit the DHR96 function and that inhibiting the function of DHR96 makes DHR96 have decreased resistance to pesticides and toxins. Also

disclosed are methods utilizing DHR96 to identify compounds that modulate its function, such as inhibit its function. Molecules that inhibit DHR96 render the insect more susceptible and sensitive to pesticides.

40. The *Drosophila* genome encodes 18 nuclear receptors that have a classical DNA-binding and ligand-binding domain and, of those, just two have identified ligands. In the nematode *C. elegans*, it was shown that a mutation in the nuclear receptor *nhr-8* gene causes a reduced resistance to colchicine and chloroquine, suggesting that this gene is involved in the xenobiotic pathway (Lindblom, T. H., et al. (2001). *Curr Biol* 11, 864-868, which is herein incorporated by reference at least for material related to nuclear receptors and their activity, and for material related to NHR8). Disclosed herein *DHR96* mutants are viable under normal conditions, but exhibit a significantly lower resistance to DDT when compared to wild type flies. Additionally, microarray analysis of animals that overexpress DHR96 indicate that this nuclear receptor regulates genes which primarily encode detoxification enzymes.

41. Disclosed herein insecticide function in insects can be reviewed from a different perspective. Disclosed are methods for identifying DHR96 antagonists and agonists. Also disclosed are methods related to the identification of the DHR96 target gene network. Also disclosed is a class of pesticides that targets the regulatory pathways that control the detoxification machinery.

#### (a) *Classes of nuclear receptors*

42. Retinoid, vitamin D, steroid, and thyroid hormones are small hydrophobic ligands that initiate a diverse array of developmental and metabolic responses. The receptors that mediate these responses form the basis of the nuclear hormone receptor superfamily (see Tsai, M.-J. & O'Malley, B. W. (1994). *Annu. Rev. Biochem.* 63, 451-486, for a review). This family is defined by a characteristic protein domain structure including a conserved DNA-binding domain and a ligand binding/dimerization domain. Members of this superfamily can be divided into three classes based on their ligand-binding and DNA-binding properties. Steroid receptors, including the estrogen and glucocorticoid receptors, form homodimers that bind to an inverted repeat of 6 bp consensus half-sites (Tsai, M.-J. & O'Malley, B. W. (1994). *Annu. Rev. Biochem.* 63, 451-486, Gronemeyer, H. (1992). *FASEB J.* 6, 2524-2529). The second class includes the retinoid receptors, RAR and RXR, as well as receptors for thyroid hormone and vitamin D. These receptors can bind to direct repeats of AGGTCA half-sites as homodimers or heterodimers (Stunnenberg, H. G. (1993). *BioEssays* 15, 309-315). The third and largest class are referred to as orphan receptors since their potential ligands are unknown. At least some of these receptors,

including Rev-Erb and NGF1-B, can bind to a single AGGTCA half-site (Harding, H. P. & Lazar, M. A. (1993). *Mol. Cell. Biol.* 13, 3113-3121; Wilson, T. E., et al., (1993). *Mol. Cell. Bio.* 13, 5794-5804). Although extensive studies have provided significant insights into the mechanisms by which nuclear hormone receptors regulate the transcription of target genes, we still know little about how these changes in gene expression result in specific and diverse developmental responses.

**(b) *Drosophila* nuclear receptors**

43. There are 18 canonical nuclear receptor genes in the complete genome of the fly *Drosophila melanogaster* (Adams et al., (2000) *Science* 287, 2185-2195, which is herein incorporated by reference at least for material related to nuclear receptors). The 18 members of the nuclear hormone receptor superfamily identified in *Drosophila* are: *EcR*, *usp*, *tll* (Pignoni, F., et al., (1990). *Cell* 62, 151-163), *svp* (Mlodzik, M., et al., (1990). *Cell* 60, 211-224), *dHNF-4* (Zhong, W., et al., (1993). *EMBO J* 12, 537-544), *E75* (Segraves, W. A. & Hogness, D. S. (1990). *Genes Dev.* 4, 204-219), *E78* (Stone, B. L. & Thummel, C. S. (1993). *Cell* 75, 307-320), *FTZ-F1* (Lavorgna, G., et al., (1991). *Science* 252, 848-851), *DHR3* (Koelle, M. R., et al., (1992). *Proc. Natl. Acad. Sci. USA* 89, 6167-6171), *DHR4* (Weller J, Sun GC, Zhou B, Lan Q, Hiruma K, Riddiford LM. Isolation and developmental expression of two nuclear receptors, MHR4 and betaFTZ-F1, in the tobacco hornworm, *Manduca sexta*. *Insect Biochem Mol Biol.* 2001 Jun 22;31(8):827-37.; King-Jones, K. Charles, J.-P., & C.S. Thummel, The DHR4 orphan nuclear receptor is required for *Drosophila* growth and metamorphosis, manuscript in prep; Adams et al., (2000) *Science* 287, 2185-2195) and *DHR39* (Ohno, C. K. & Petkovich, M. (1992). *Mech. Dev.* 40, 13-24; Ayer, S., et al., (1993). *Nuc. Acids Res.* 21, 1619-1627), *DHR38*, *DHR78* (Fisk and Thummel, (1995), PNAS, Proc Natl Acad Sci U S A. 1995 Nov 7;92(23):10604-8), *DHR83* (King-Jones, K. and C.S. Thummel (2003) *Drosophila* nuclear receptors. In "Handbook of Cell Signaling," Vol. 3, (Bradshaw, R. and Dennis, E., eds.), Academic Press, New York, pp. 69-73; Adams et al., (2000) *Science* 287, 2185-2195), *DHR96* (Fisk and Thummel, 1993), *dsf* (Finley, K. D., et al. (1998). "dissatisfaction encodes a Tailless-like nuclear receptor expressed in a subset of CNS neurons controlling *Drosophila* sexual behavior." *Neuron* 21, 1363-1374), *dERR* (King-Jones, K. and C.S. Thummel (2003) *Drosophila* nuclear receptors. In "Handbook of Cell Signaling," Vol. 3, (Bradshaw, R. and Dennis, E., eds.), Academic Press, New York, pp. 69-73; Adams et al., (2000) *Science* 287, 2185-2195), and *dFAX-1* (King-Jones, K. and C.S. Thummel (2003) *Drosophila* nuclear receptors. In "Handbook of Cell Signaling," Vol. 3, (Bradshaw, R. and Dennis, E., eds.), Academic Press, New York, pp. 69-73; Adams et al., (2000) *Science* 287,

2185-2195) At least seven of these genes appear to contribute to the 20E regulatory hierarchies that direct the onset of metamorphosis – *E75*, *E78*, *βFTZ-F1*, *DHR3*, *DHR39*, *EcR*, and *usp* (Richards, G. (1992). *Current Biology* 2, 657-659; Horner, M., et al., (1995). *Dev. Biol.* 168, 490-502; Woodard, C. T., et al., (1994). *Cell* 79, 607-615).

5 44. Table 5 provides a list of *Drosophila* nuclear receptors.

45. Table 5

probe set	CG	CT	Accession	Description sym=Hr4 orEG:133E12.2	SEQ ID NO
144004_at	CG16902	CT37504	FBgn0023546	/name= DHR4	SEQ ID NO:1
154699_at	CG4059	CT13432	FBgn0001078	sym=ftz-f1 /name=ftz transcription factor 1	SEQ ID NO:3
143123_at	CG11823	CT11367	FBgn0000448	sym=Hr46 or DHR3 /name=Hormone receptor-like in 46	SEQ ID NO: 5
152580_at	CG11783	CT33046	FBgn0015240	sym=Hr96 or DHR96/name=Hormone receptor-like in 96	SEQ ID NO: 7
143535_at	CG9310	CT40906	FBgn0004914	sym=Hnf4 /name=Hepatocyte nuclear factor 4	SEQ ID NO: 9
143768_at	CG1864	CT5732	FBgn0014859	sym=Hr38 or DHR38 /name=Hormone receptor-like in 38	SEQ ID NO: 11
149398_at	CG10296	CT28911	FBgn0037436	sym=CG10296 or DHR83 /name=Hr83	SEQ ID NO: 13
143372_at	CG11502	CT12919	FBgn0003651	sym=svp /name=seven up /prod=nuclear receptor NR2F3	SEQ ID NO: 15
143379_at	CG1378	CT3134	FBgn0003720	sym=tll /name=tailless /prod=nuclear receptor NR2E2	SEQ ID NO: 17
143805_at	CG9019	CT25922	FBgn0015381	sym=dsf /name=dissatisfaction /prod= /func=receptor	SEQ ID NO: 19
147244_at	CG16801	CT37351	FBgn0034012	sym=CG16801 /name=FAX-1 /prod=nuclear hormone receptor-like	SEQ ID NO: 21
153072_at	CG7404	CT22787	FBgn0035849	sym=CG7404 /name=ERR /prod= /func=steroid hormone receptor	SEQ ID NO: 23
152160_at	CG7199	CT22217	FBgn0015239	sym=Hr78 or DHR78/name=Hormone- receptor-like in 78	SEQ ID NO: 25
153675_at	CG4380	CT14272	FBgn0003964	sym=usp /name=ultraspiracle /prod=nuclear receptor NR2B4	SEQ ID NO: 27
153197_at	CG8127	CT24290	FBgn0000568	sym=Eip75B or E75/name=Ecdysone-induced protein 75B	SEQ ID NO: 29
143525_at	CG18023	CT40336	FBgn0004865	sym=Eip78C or E78/name=Ecdysone-induced protein 78C	SEQ ID NO: 31

154377_at	CG1765	CT5200	FBgn0000546	sym=EcR /name=Ecdysone receptor /prod=ecdysone receptor	SEQ ID NO: 33
155094_at	CG8676	CT5296	FBgn0010229	sym=EcR /name=Ecdysone receptor /prod=ecdysone receptor	SEQ ID NO: 35
46.					

47. While there are 18 nuclear receptors in flies, there are 48 in humans (Robinson-Rechavi et al., (2001) *Trends Genet* 17, 554-556), 49 in the mouse with the addition of FXR $\beta$ , (Robinson-Rechavi and Laudet, 2003, *Methods Enzymol.* 2003;364:95-118) and more than 270 genes in the nematode worm *Caenorhabditis elegans* (Sluder et al., (1999). *Genome Research* 9, 103-120.

### (c) *Role of 20-hydroxyecdysone(20E) in Drosophila*

48. 20E is involved in the metamorphosis of the fruit fly, *Drosophila melanogaster* through steroid hormone receptors. A high titer 20E pulse at the end of third instar larval development triggers puparium formation, followed 10 hrs later by an 20E pulse that triggers head eversion and the onset of pupal development (Pak, M. D. & Gilbert, L. I. (1987). *J. Liq. Chrom.* 10, 2591-2611; Richards, G. (1981). *Mol. Cell. Endocrin.* 21, 181-197). The 20E receptor is encoded by two members of the nuclear hormone receptor superfamily, *EcR* (Koelle, M. R., et al., (1991). *Cell* 67, 59-77) and *usp* (Henrich, V. C., et al., (1990). *Nuc. Acids Res.* 18, 4143-4148; Shea, M. J., et al., (1990). *Genes Dev.* 4, 1128-1140; Oro, A. E., et al., (1990). *Nature* 347, 298-301). *Usp* is most closely related to the vertebrate RXR family and can heterodimerize with vertebrate thyroid and vitamin D receptors, as well as with *EcR* (Yao, T., et al., (1992). *Cell* 71, 63-72; Thomas, H. E., et al., (1993). *Nature* 362, 471-475; Yao, T., et al., (1993). *Nature* 366, 476-479; Koelle, M. R. (1992) Ph.D. thesis, Stanford University). The ability of RXRs to function as promiscuous heterodimerization partners combined with the sequence similarity of many receptor binding sites raises the possibility that other members of the superfamily may function in transducing 20E signals, either by interacting directly with *EcR* and/or *Usp*, or by competing for receptor binding sites (Richards, G. (1992). *Current Biology* 2, 657-659).

### (d) *General structure of nuclear receptors*

49. There are a number of domains in a nuclear receptor. From the N terminus to the C terminus there is the A/B domain, followed by a DNA binding domain (DBD, C), which contains the DNA sequence recognition domain called the P-box, which is followed by a less conserved region, D, which acts as a flexible hinge between the DBD and the ligand binding domain (LBD, E) and the D domain typically contains the nuclear localization signal, but this

may overlap with the C domain, and finally some nuclear receptors contain a C-terminal F domain whose function is unknown.

50. The A/B domain and N terminal region in general is highly variable and can range in size from less than about 50 amino acids to more than about 500 amino acids. The A/B domain typically contains the transactivation domains which typically include at least one constitutively active domain, the AF-1 domain, and then typically one or more autonomous activation domains which can be regulated or not, called AD domains.

51. The DBD is typically the most conserved region. It contains the P-box, a six amino acid region that confers specificity for binding to particular target sites in the DNA. The P-box for DHR96 is ESCKA. An example of DHR96 is shown in SEQ ID NO:7. The DBD is also typically the site of homo- and hetero- dimerization. The 3D structure of the DBD shows that it contains contains two highly conserved zinc- fingers – C-X2-C-X13-C-X2-C and CX5- C-X9-C-X2-C – the four cysteines of each finger chelating one Zn<sup>2+</sup> ion.

52. The LBD is typically the largest domain and is only moderately conserved, but the secondary structure is often conserved and contains 12  $\alpha$ -helices. Many functions are associated with the E domain, including the AF-2 transactivation function, a strong dimerization interface, another NLS, and often a repression function. Typically the functions are ligand regulated.

#### ***(e) Dimerization of nuclear receptors.***

53. Dimerization of nuclear receptors is very important to their function. The dimerization domains typically reside in the DBD and LBD. Many nuclear receptors heterodimerize with RXRs (USP in arthropods), such as DHR38 (NR4A4), NGFIB (NR4A1), NURR1 (NR4A2), NOR1 (NR4A3), LXR and FXR subfamilies (LXR $\alpha$  (NR1H3), LXR $\beta$  (NR1H2, HO), ECR (NR1H1), FXR $\alpha$  (NR1H4, HO), FXR $\beta$  (NR1H5, HO), the CAR1 and VDR subfamilies including, CAR1 (NR1I3), PXR (NR1I2), VDR (NR1L1) (NR1J1), the PPAR subfamily including, PPAR $\gamma$  (NR1C3), PPAR $\alpha$  (NR1C1), AND PPAR $\beta$  (NR1C2), the RAR subfamily including RAR $\beta$  (NR1B2), RAR $\alpha$  (NR1B1), and RAR $\gamma$  (NR1B3), and TR $\alpha$  (NR1A1), and TR $\beta$  (NR1A2), and possibly COUP-TF and FXR $\beta$  (for a review, see Robinson-Rechavi M, Escriva Garcia H, Laudet V., J Cell Sci. 2003 Feb 15;116(Pt 4):585-6). DHR96 can also be found to dimerize with any other receptor, such as USB, or itself.

#### ***(f) Ligands for nuclear receptors***

54. The superfamily includes receptors for many different types of molecules. For example, nuclear receptors bind hydrophobic molecules such as steroid hormones, such as estrogens, glucocorticoids, progesterone, mineralocorticoids, androgens, vitamin D3, ecdysone,

oxysterols and bile acids. Certain nuclear receptors also bind retinoic acids, such as all-trans and 9-cis isoforms, thyroid hormones, fatty acids, leukotrienes and prostaglandins (Escriva et al., 2000, Bioessays 22, 717-727 and Robinson-Rechavi M, Escriva Garcia H, Laudet V., J Cell Sci. 2003 Feb 15;116(Pt 4):585-6).

5

**(g) How nuclear receptors function**

55. Nuclear receptors typically act in a stepwise fashion that starts with repression, moves to a state of derepression, and ends with transcription activation. (reviewed by Robinson-Rechavi M, Escriva Garcia H, Laudet V., J Cell Sci. 2003 Feb 15;116(Pt 4):585-6).

56. Repression typically occurs with corepressors, such as the histone deacetylase activity (HDAC) (for example, the apo-nuclear receptor). Usually ligand binding results in derepression, caused by the disassociation of the receptor from the corepressors. Also ligand binding typically causes the recruitment of coactivators, such as histone acetyltransferase (HAT) activity, which causes chromatin decondensation, which is believed to be necessary but not sufficient for activation of the target gene. After the HAT complex dissociates, typically a second coactivator complex is assembled (TRAP/DRIP/ARC), which is able to establish contact with the basal transcription machinery, and thus results in transcription activation of the target gene, but many other transcription co-activators can be associated with the nuclear receptor and these coactivators can provide activation discrimination. This general scheme does not apply for all nuclear receptors, as for example, some nuclear receptors can activate without ligand and some may bind DNA without ligand and some may repress with or without ligand.

**(2) DHR96 gene**

57. *DHR96* maps to 96B12-14 in the polytene chromosomes of *Drosophila*. The DHR96 gene was cloned and sequenced and its sequence is set forth in SEQ ID NO:1. (Fisk and Thummel (1995) Proc. Natl. Acad. Sci USA, 92: 10604-10608, herein incorporated by reference at least for material related to the DHR96 gene and its sequence including the specific sequence).

58. *DHR96* is highly conserved in *Anopheles gambiae*, a distant (~ 250 M years) dipteran species (see Table 4). Similarly, many other *Drosophila* nuclear receptors are conserved in even more distant insects and, when examined, their regulatory functions appear to be conserved as well (Swevers L, Iatrou K. The ecdysone regulatory cascade and ovarian development in lepidopteran insects: insights from the silkworm paradigm. Insect Biochem Mol Biol. 2003 Dec;33(12):1285-97; Riddiford LM, Hiruma K, Zhou X, Nelson CA. Insights into the molecular basis of the hormonal control of molting and metamorphosis from *Manduca sexta* and *Drosophila melanogaster*. Insect Biochem Mol Biol. 2003 Dec;33(12):1327-38). This is

consistent with the role of detoxification via *DHR96* being conserved through evolution. Thus, inactivation of *DHR96* function in known insect pests provides a novel mode of intervention. It is understood that *DHR96* homologs in other insects, insect orders, insect families and other insect species are considered disclosed and that they function in a manner similar to *DHR96* in *Drosophila*. There is significant homology within the order Diptera and within the class of insects in general for nuclear receptors, and there is shown in Table 4, that there is a high degree of homology between *DHR96* in other insects, such as the mosquito.

59. Disclosed are *DHR96* variants that have at least 60%, 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity or homology as discussed herein in to the LBD of *DHR96*, DBD of *DHR96*, or full length *DHR96*, or of fragments of *DHR96*, functional or otherwise.

60. Among the *C. elegans* receptors, *DHR96* is most similar to *DAF-12*, which is a gene involved in dauer larva formation in *C. elegans* (68% identity DBD; 29% identity LBD). The match with *NHR-8* in *C. elegans* is weaker (60%; 25%). This is consistent with *DHR96* having a role similar to *DAF-12*. *DAF-12* reads signals from *TGFbeta* and insulin and decides when the worm should enter diapause to survive difficult conditions. Diapause is similar to pupal stages in many ways (indeed many insects diapause during metamorphosis). Disclosed herein, mutants of *DHR96* did not have any effects on metamorphosis – and they survived. Thus it was expected that *DHR96* would have a function similar to *DAF-12*. *DAF-12* is a gene involved in dauer larva formation in *C. elegans*. *DAF-12* reads signals from *TGFbeta* and insulin and decides when the worm should enter diapause to survive difficult conditions. Diapause is similar to pupal stages in many ways (indeed many insects diapause during metamorphosis). However, as disclosed herein, mutants of *DHR96* did not have any effects on metamorphosis – as they survived.

61. Disclosed are systems that assay for effects of drugs that alter *DHR96* – and thus one can assay for effects on target gene transcription and relate that expression to the ability of an animal, such as an insect, to resist toxins.

#### 62. Table 4

species	DBD amino acids 7-72 identity	p-box	LBD amino acids 501-723 identity
<i>anopholes gambiae</i>	86% %	same	65% %
<i>c.elegans daf-12</i>	69%	same	26%
<i>strongyloides stercoralis-parasitic worm</i>	67%	different	27%
<i>c.elegans nhr-48</i>	66%	same	

	%		
VDR-zebrafish	65%	different	27%
VDR-bastard halibut	63%	different	27%
mouse vdr	62%	different	23%
human vdr	62%	different	24%
c.elegans nhr-8	60%	same	25%
mouse pxx	59%	different	23%
human pxx	59%	different	22%
human car	56%	different	19%
AamEcRA1-tick	54%	different	
ecdysone receptor-locusta			
migratoris-locust	53%	different	
ecdysone receptor-calliphor vicina-			
insect	53%	different	
EcR- tenebrio molitor-yellow			
mealworm	53%	different	
EcR- d. melanogaster	51%	different	
EcR- aedes albopictus-mosquito	51%	different	
mouse car	51%	different	20%

63.

64. Table 4 shows the percent identical amino acids within the DNA binding domain and ligand binding domain for DHR96 and the best matches in the public databases (Genbank).

Shown is the mosquito DHR96 gene, and it is the orthologous receptor in mosquito. (anopholes gambiae) (85% and 65% identity - very high). Also listed is whether the sequence within the P box, is either the same as DHR96 or different. This sequence directs the DNA binding specificity of the receptor. DHR96 DNA binding is predicted to be similar to that of all three nematode homologs (daf-12, nhr-48 and nhr-8), but none of the vertebrate ones.

65. In certain embodiments homologs of DHR96 in other insect species can have at least 50% identity in the DBD and 25% identity in the LBD.

66. An alignment of the *Drosophila* nuclear hormone receptor DNA-binding domains reveals a central region of 8-9 unique amino acids flanked by highly conserved regions that each contain a C<sub>2</sub>C<sub>2</sub> zinc finger (Fig. 5).

67. The DNA-binding domain of DHR96 is 64% identical to the human vitamin D receptor and 52% identical to EcR (Fig. 6C). The DHR96 ligand binding domain (amino acids 501 - 723) is most similar to that of thyroid hormone receptor, with 23% identity.

68. *DHR96* encodes a 2.8 kb transcript that is expressed throughout third instar larval and prepupal development, with distinct increases in abundance at 106 hrs after egg laying (Fig. 7). The temporal patterns of *DHR96* transcription most closely resemble those of the genes encoding the 20E receptor. *EcR* and *usp* mRNAs can be detected throughout third instar larval and

prepupal development (Andres, A. J., et al., (1993). *Dev. Biol.* **160**, 388-404; 36; Henrich, V. C., et al., (1994). *Dev. Biol.* **165**, 38-52).

69. The *hsp27* EcRE is the only oligonucleotide bound by DHR96, albeit it a weak interaction (Fig. 9). The EcRE consists of a palindromic arrangement of the imperfect half-sites AGtgCA and gGtTCA. DHR78 and DHR96 recognize distinct sequences that can also be bound by the EcR/Usp heterodimer (Horner, M., et al., (1995). *Dev. Biol.* **168**, 490-502). These distinct binding specificities are consistent with the P-box sequences of the DHR78 and DHR96 proteins. The DHR78 P-box, EGCKG, like that of DHR38, directs binding to an AGGTCA half-site sequence (Tsai, M.-J. & O'Malley, B. W. (1994). *Annu. Rev. Biochem.* **63**, 451-486). In contrast, DHR96 contains a unique P-box sequence that is only present in its three *C. elegans* homologs (see Table 4 above) – ESCKA. The binding of the *hsp27* EcRE by DHR96 is very weak. An optimal DNA binding site can be identified by further experimentation.

70. It will be of interest to determine whether DHR78 or DHR96 can heterodimerize with EcR, Usp, or any of the *Drosophila* orphan receptors.

#### (a) *DHR96 functions in the xenobiotic pathway*

71. Several lines of evidence support the conclusion that *DHR96* acts in a xenobiotic pathway. First, the protein is selectively expressed in tissues involved in nutrient absorption (gastric caecae), metabolism (fat body), and excretion (Malpighian tubules) – tissues that should play a primary role in detoxification and elimination of both endobiotic and xenobiotic compounds. Second, *DHR96* mutants, like null mutants in the mouse PXR and CAR xenobiotic nuclear receptors, are viable and fertile, indicating no critical role in normal development. Third, *DHR96* mutants are more sensitive to the pesticide DDT. Fourth, the most highly repressed genes in response to *DHR96* overexpression comprise members of all four classes of insect detoxifying genes.

72. The effect of the mutants can be confirmed by the expression of wild type *DHR96* (from a heat-inducible *DHR96* transgene, for example) in a homozygous mutant background, and test for DDT sensitivity. This experiment should rescue the sensitivity back to wild type levels. In addition, *DHR96* function was reduced by RNAi and this results in levels of DDT sensitivity that are similar to those of *DHR96* mutants.

73. The decreased resistance to DDT in *DHR96* mutants can be confirmed as related to the inability to neutralize toxins rather than a general lack of fitness by demonstrating that sensitivity of *DHR96* mutants occurs for toxic compounds. It can also be confirmed by showing that detoxifying genes fail to be induced in *DHR96* mutants treated with toxic compounds, by for

example, microarray analysis, with the mutants in the presence or absence of a toxin. These results could be compared to the microarray data disclosed herein. Two toxins that could be used for this are DDT and phenobarbital because the latter was shown to induce a number of cytochrome P450 genes in *Drosophila* (Danielson, P. B. et al. (1998) Mol Gen Genet 259, 54-59).

74. The expression of DHR96 and its activation level can be assayed to determine if it is directly activated by toxic compounds, similar to the ability of xenobiotics to bind to human PXR xenobiotic nuclear receptor. This can be done using transformed *Drosophila* that express a fusion of the yeast GAL4 DNA binding domain to the ligand binding domain of DHR96. When combined with a GAL4-dependent *lacZ* reporter gene, the expression of  $\beta$ -galactosidase will only occur when the DHR96 ligand binding domain is in an active conformation. This could be caused by a direct interaction between DHR96 and the toxin. Larval organs that carry these constructs can be cultured in the presence of various xenobiotic inducers, testing for induction of *lacZ* reporter gene activity. Furthermore, target gene promoters can be identified which can also demonstrate a direct interaction between DHR96 and the expression of a detoxifying enzyme.

75. In the disclosed microarray study, *DHR96* was overexpressed and it was found that this resulted in repression of a significant number of members of the major detoxification gene families. Repression of cuticle proteins was also observed, consistent with a role for cuticle formation in inhibiting pesticide toxicity (Wilson, T. G. (2001). Annu Rev Entomol 46, 545-571). The observation that these target genes are repressed suggests that DHR96 might function as a repressor in the absence of ligand. This is consistent with the action of other nuclear receptors, for example, both Endocrine receptor (EcR) and thyroid receptor (TR) are known to function in this manner. Very strict filtering criteria were used in the disclosed microarray experiments further strengthening the results.

76. The microarray studies allow the identification of the direct targets of DHR96. This will allow the identification of the genetic hierarchy that is regulated by this nuclear receptor. Once target genes have been identified, it will be possible to construct reporter genes that are inducible by endogenous DHR96. Such a system can then be utilized to screen for drugs or combinations of drugs that activate or repress these reporter genes, in both a wild type and *DHR96* mutant background. This can further confirm that *DHR96* can directly regulate the expression of detoxifying genes. This system would also provide a direct readout of DHR96 activity that would be useful for further studies of *DHR96* function and for the development of appropriate inhibitors of DHR96 function. The mutants of DHR96 can be used to identify and

confirm other factors that can act as xenobiotic receptors in insects, and test whether these act in a partially redundant manner with *DHR96*.

77. As disclosed herein, PXR and DHR96 are highly homologous. PXR transactivation and binding assays have been developed into high-throughput assays (Zhu et al., J Biomol Screen. 2004 Sep;9(6):533-40; Kliewer et al., Endocrine Rev. 2002 23(5):687-702 herein incorporated by reference in its entirety for its teaching concerning PXR, transactivation assays, and binding assays.) Zhu et al. found a good correlation between the results of the transactivation and binding assays. An example of an antagonist of PXR is ecteinascidin-743. Furthermore, several compounds can activate DHR96, such as tebufenozide (RH-5992, Fig. 13) (Dinan et al. 1997 Biochem J. 327:643-50,). This compound is both an ecdysteroid agonist and a lepidopteran insecticide.

78. The steroid and xenobiotic receptor (SXR) is another nuclear receptor with a high degree of homology with DHR96. SXR is a nuclear receptor that regulates drug clearance in the liver and intestine via induction of genes involved in drug and xenobiotic metabolism. The  $\alpha$ ,  $\beta$ ,  $\Delta$ , and  $\gamma$  tocotrienols specifically bind to and activate SXR (Zhou et al. Drug Metab Dispos. 2004 Oct;32(10):1075-82, herein incorporated by reference for its teaching concerning SXR). Many other compounds also activate SXR and can be activators of DHR96 as well (Blumberg et al. Genes Dev. 1998 Oct 15 12(20):3195-205, herein incorporated by reference in its entirety for its teaching regarding nuclear receptor modulators.)

79. Nuclear receptors, such as DHR96, SXR, and PXR, contain a lipophilic ligand binding pocket. This pocket can be bound by compounds that affect the activity of the nuclear receptor, and therefore act as selective modulators of the nuclear receptor. These selective modulators can act as either agonists or antagonists, and modulators of one nuclear receptor can act as modulators of another.

### (3) Mutants of the DHR96 gene

80. Various DHR96 mutant alleles were made. A series of studies to characterize the *DHR96* mutant alleles were performed. These included Southern, Northern and Western blotting, tissue stains, sequencing of PCR products, and genetic mapping to validate the mutations in the different *DHR96* alleles. Validation of these alleles was particularly important because flies homozygous for *DHR96* mutations are viable and fertile. At least one of the alleles generated, *DHR96*<sup>16A</sup>, is a protein null, because the translation start site was deleted and no protein was detectable in Western blots or tissue stains of homozygous mutant animals.

81. Gene targeting (Rong, Y. S., and Golic, K. G. (2000). Science 288, 2013-2018) was used to generate mutations in *DHR96* because no deficiencies or P elements were known in this region of the genome. (see Example 1). Using these methods any mutations of the *DHR96* gene can be made, such as mutations at or around the start site; mutations at or around the splice sites; mutations which prevent or render inactive complete or partial exon sequences; mutations which render inactive or remove the complete or partial DBD or LBD or any of the domains of *DHR96* discussed herein that it contains as a nuclear receptor.

82. The *DHR96* gene resides on the third chromosome. When mutations are made in certain embodiments the mutations of the *DHR96* gene are made such that there is only a single copy of the mutant and no copies of the wildtype gene in the insect, such as the fly. This is done, for example, by using vectors for the mutation generation, which have sites built in that allow for recombination and excision of the site, and fly stocks containing a single copy can be selected. (see for example, Rong, Y. et al., (2002) Genes Dev 16, 1568-1581).

83. Disclosed are null mutants of the *DHR96* gene. A null mutant is defined herein as a mutant that lacks functional *DHR96* protein product.

84. A null mutant disclosed herein is *DHR96*<sup>16A</sup> which is mutant having two specific deletions, one removing the start codon for translation and the second removing intron/exon 4, deleting a critical portion of the LBD.

85. Another null mutant disclosed herein is the mutant *DHR96*<sup>E25</sup> which carries a tandem duplication of the *DHR96* gene in place of the single wild type copy. One of these mutant *DHR96* genes is identical to the *DHR96*<sup>16A</sup> allele described above, missing both the start codon and intron/exon 4. The other mutant *DHR96* gene is lacking only intron/exon 4. Western blot analysis indicates that both *DHR96*<sup>E25</sup> mutants, as well as *DHR96*<sup>16A</sup> mutants, produce no detectable *DHR96* protein. Thus, both alleles can be considered as null mutations.

86. One way to functionally test the mutants is in a viability assay based on different nutritional backgrounds. Disclosed herein, *DHR96* mutants will have a decreased ability to grow on instant fly food, such as Carolina 424. If yeast is restored to the instant food, viability is restored to within wildtype levels, indicating that *DHR96* mutants are sensitive to the absence of yeast in their food source. In contrast, mutants such as *DHR96*<sup>E25</sup> or *DHR96*<sup>16A</sup> are viable when grown on standard cornmeal medium.

87. Disclosed are insects, such as flies, containing the mutant *DHR96* gene, as well as any of their developmental stages, such as larvae, eggs, or pupae. These flies can be used, for example, to be crossed with other strains of flies to make new strains harboring the *DHR96*

mutants. These strains could also be used, for example, as a type of insect inhibitor themselves, by being released in the wild to cross with wildtype insects creating mutant insects. For this purpose, mutations that create a dominant negative phenotype are preferred, such as those that have non-functional LBD, but retain their ability to heterodimerize, thus, interacting with and  
5 reducing the effect of native proteins in the insect.

88. The disclosed mutants cause a decrease in the insect's ability to react to toxins or pesticides, such as DDT. The disclosed mutants, such as *DHR96*<sup>16A</sup> or *DHR96*<sup>E25</sup> insects, such as flies, were more sensitive to DDT and died at lower concentrations of DDT compared to control animals (Fig. 4). In addition, when challenged with a fixed concentration of DDT,  
10 *DHR96* homozygotes died more rapidly than wild type flies (Fig. 10).

89. Also disclosed are mutants which have a defect in for example, activation with and without retention of dimerization ability, defects in ligand binding, and defects in DNA binding with and without loss of dimerization ability.

90. Also disclosed are mutants that, when overexpressed, fail to modulate genes in the xenobiotic pathway, such as genes in the four major detoxification families, cytochrome P450s, carboxylesterases, glutathione S-transferases, and UDP-glucuronosyltransferases (Oakeshott JG, Home I, Sutherland TD, Russell RJ. The genomics of insecticide resistance. *Genome Biol.* 2003;4(1):202). In Table 3, two are P450s (Cyp genes), two are glutathione S-transferases, and one each of the carboxylesterases and UDP-glucuronosyltransferases were identified by  
15 microarray analysis. These represent the function of these proteins. Also denoted in Table 3 are the names of the genes. These are the gene names according to FlyBase (<http://flybase.bio.indiana.edu/>) They are either a proper name, like black or Lcp1, or the CG number, which is a numerical designation given to each fly gene. The CG number is usually used when the gene is new or of unknown function. This can be determined using microarrays as  
20 disclosed herein.  
25

#### (4) Compounds that modulate DHR96 activity

91. Disclosed are compounds that modulate DHR96 activity. These compounds can, for example, modulate the activity of the protein through binding with the protein of DHR96, or through binding the mRNA of DHR96, and inhibiting the mRNA, through, for example,  
30 degradation or prevention of translation. The compositions can be any type of molecule, including, for example, proteins, small peptides, antibodies, functional nucleic acids, such as aptamers, antisense, ribozymes, dsRNA for RNAi or siRNA, or small molecules, such as those found in various combinatorial chemistry libraries or natural product libraries.

92. For example, disclosed are compounds that function by, for example, binding to the ligand binding domain of DHR96 and inactivating its function or turning it into a constitutive repressor, or mimicking the normal cofactors that mediate nuclear receptor signaling to the general transcription machinery. These compounds, such as peptides, would render the receptor incapable of directing proper target gene transcription, blocking the detoxification response. The disclosed compounds can act in combination with known or any pesticide by increasing the effectiveness of the pesticide by decreasing the insect's ability to react to the pesticide. The compositions could be added to pre-existing pesticide formulations, increasing their effectiveness. Moreover, resistant lines of insects that respond poorly to a particular pesticide may be made more sensitive by adding compounds that affect DHR96 function. DHR96 is a target for pest control, capable of regulating insect populations. The compositions could also prevent or reduce the translation or expression of the DHR96 mRNA, by for example, through RNAi or antisense mechanisms.

*(a) Functional Nucleic Acids*

93. Functional nucleic acids are nucleic acid molecules that have a specific function, such as binding a target molecule or catalyzing a specific reaction. Functional nucleic acid molecules can be divided into the following categories, which are not meant to be limiting. For example, functional nucleic acids include RNAi, antisense molecules, aptamers, ribozymes, triplex forming molecules, and external guide sequences. The functional nucleic acid molecules can act as effectors, inhibitors, modulators, and stimulators of a specific activity possessed by a target molecule, or the functional nucleic acid molecules can possess a de novo activity independent of any other molecules.

94. Functional nucleic acid molecules can interact with any macromolecule, such as DNA, RNA, polypeptides, or carbohydrate chains. Thus, functional nucleic acids can interact with the mRNA of DHR96 or variants or fragments or the genomic DNA of DHR96 or variants or fragments or they can interact with the polypeptide DHR96 or variants or fragments. Often functional nucleic acids are designed to interact with other nucleic acids based on sequence homology between the target molecule and the functional nucleic acid molecule. In other situations, the specific recognition between the functional nucleic acid molecule and the target molecule is not based on sequence homology between the functional nucleic acid molecule and the target molecule, but rather is based on the formation of tertiary structure that allows specific recognition to take place.

95. Disclosed are molecules that inhibit DHR96 activity that are based on RNA interference (RNAi) or small interfering RNA (SiRNA). It is thought that RNAi involves a two-step mechanism for RNA interference (RNAi): an initiation step and an effector step. For example, in the first step, input double-stranded (ds) RNA is processed into small fragments (siRNA), such as 21–23-nucleotide 'guide sequences'. RNA amplification appears to be able to occur in whole animals. Typically then, the guide RNAs can be incorporated into a protein RNA complex which is cable of degrading RNA, the nuclease complex, which has been called the RNA-induced silencing complex (RISC). This RISC complex acts in the second effector step to destroy mRNAs that are recognized by the guide RNAs through base-pairing interactions. RNAi involves the introduction by any means of double stranded RNA into the cell which triggers events that cause the degradation of a target RNA. RNAi is a form of post-transcriptional gene silencing. Disclosed are RNA hairpins that can act in RNAi.

96. RNAi has been shown to work in a number of cells, including mammalian and invertebrate cells. In certain embodiments the RNA molecules which will be used as targeting sequences within the RISC complex are shorter. For example, less than or equal to 50 or 40 or 30 or 29, 28, 27, 26, 25, 24, 23, 22, 21, 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, or 10 nucleotides in length. These RNA molecules can also have overhangs on the 3' or 5' ends relative to the target RNA which is to be cleaved. These overhangs can be at least or less than or equal to 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, or 20 nucleotides long.

97. Methods of RNAi and SiRNA are described in detail in Hannon et al. (2002), RNA Interference, Nature 418:244-250; Brummelkamp et al. (2002), A System for Stable Expression of Short Interfering RNAs in Mammalian Cells, Science 296:550-508; Paul et al. (2002), Effective expression of small interfering RNA in human cells, Nature Biotechnology 20: 505-508, which are each incorporated by reference in their entirety for methods of RNAi and SiRNA and for designing and testing various oligos useful therein.

98. RNA interference (RNAi) and gene targeting were used to disrupt *DHR96* function because no existing mutants were available. The effects of *DHR96* RNAi were analyzed by generating transgenic lines that express snapback RNA under the control of a heat-inducible promoter. Three independent lines showed strong reduction of *DHR96* mRNA in northern blots when treated with a single heat-shock, but displayed no discernable phenotype. Using a variety of heat-shock regimens, e.g. longer single and double treatments or 12 hr repetitions, did not affect the outcome of this observation. These findings suggest that *DHR96* mRNA is not

necessary for viability under standard conditions, indicating either that DHR96 protein is very stable or dispensable for survival, and is consistent with the studies of *DHR96* null mutants.

99. Antisense molecules are designed to interact with a target nucleic acid molecule through either canonical or non-canonical base pairing. The interaction of the antisense molecule and the target molecule is designed to promote the destruction of the target molecule through, for example, RNaseH mediated RNA-DNA hybrid degradation. Alternatively the antisense molecule is designed to interrupt a processing function that normally would take place on the target molecule, such as transcription or replication. Antisense molecules can be designed based on the sequence of the target molecule. Numerous methods for optimization of antisense efficiency by finding the most accessible regions of the target molecule exist. Exemplary methods would be in vitro selection experiments and DNA modification studies using DMS and DEPC. It is preferred that antisense molecules bind the target molecule with a dissociation constant ( $k_d$ ) less than or equal to  $10^{-6}$ ,  $10^{-8}$ ,  $10^{-10}$ , or  $10^{-12}$ . A representative sample of methods and techniques which aid in the design and use of antisense molecules can be found in the following non-limiting list of United States patents: 5,135,917, 5,294,533, 5,627,158, 5,641,754, 5,691,317, 5,780,607, 5,786,138, 5,849,903, 5,856,103, 5,919,772, 5,955,590, 5,990,088, 5,994,320, 5,998,602, 6,005,095, 6,007,995, 6,013,522, 6,017,898, 6,018,042, 6,025,198, 6,033,910, 6,040,296, 6,046,004, 6,046,319, and 6,057,437.

100. Aptamers are molecules that interact with a target molecule, preferably in a specific way. Typically aptamers are small nucleic acids ranging from 15-50 bases in length that fold into defined secondary and tertiary structures, such as stem-loops or G-quartets. Aptamers can bind small molecules, such as ATP (United States patent 5,631,146) and theophiline (United States patent 5,580,737), as well as large molecules, such as reverse transcriptase (United States patent 5,786,462) and thrombin (United States patent 5,543,293). Aptamers can bind very tightly with  $k_d$ s from the target molecule of less than  $10^{-12}$  M. It is preferred that the aptamers bind the target molecule with a  $k_d$  less than  $10^{-6}$ ,  $10^{-8}$ ,  $10^{-10}$ , or  $10^{-12}$ . Aptamers can bind the target molecule with a very high degree of specificity. For example, aptamers have been isolated that have greater than a 10000 fold difference in binding affinities between the target molecule and another molecule that differ at only a single position on the molecule (United States patent 5,543,293). It is preferred that the aptamer have a  $k_d$  with the target molecule at least 10, 100, 1000, 10,000, or 100,000 fold lower than the  $k_d$  with a background binding molecule. It is preferred when doing the comparison for a polypeptide for example, that the background molecule be a different polypeptide. For example, when determining the specificity of aptamers

to DHR96 protein or fragments or variants, the background protein could be serum albumin. Representative examples of how to make and use aptamers to bind a variety of different target molecules can be found in the following non-limiting list of United States patents: 5,476,766, 5,503,978, 5,631,146, 5,731,424 , 5,780,228, 5,792,613, 5,795,721, 5,846,713, 5,858,660 ,  
5 5,861,254, 5,864,026, 5,869,641, 5,958,691, 6,001,988, 6,011,020, 6,013,443, 6,020,130, 6,028,186, 6,030,776, and 6,051,698.

101. Ribozymes are nucleic acid molecules that are capable of catalyzing a chemical reaction, either intramolecularly or intermolecularly. Ribozymes are thus catalytic nucleic acid. It is preferred that the ribozymes catalyze intermolecular reactions. There are a number of  
10 different types of ribozymes that catalyze nuclease or nucleic acid polymerase type reactions which are based on ribozymes found in natural systems, such as hammerhead ribozymes, (for example, but not limited to the following United States patents: 5,334,711, 5,436,330, 5,616,466, 5,633,133, 5,646,020, 5,652,094, 5,712,384, 5,770,715, 5,856,463, 5,861,288, 5,891,683, 5,891,684, 5,985,621, 5,989,908, 5,998,193, 5,998,203, WO 9858058 by Ludwig and  
15 Sproat, WO 9858057 by Ludwig and Sproat, and WO 9718312 by Ludwig and Sproat) hairpin ribozymes (for example, but not limited to the following United States patents: 5,631,115, 5,646,031, 5,683,902, 5,712,384, 5,856,188, 5,866,701, 5,869,339, and 6,022,962), and tetrahymena ribozymes (for example, but not limited to the following United States patents: 5,595,873 and 5,652,107). There are also a number of ribozymes that are not found in natural  
20 systems, but which have been engineered to catalyze specific reactions de novo (for example, but not limited to the following United States patents: 5,580,967, 5,688,670, 5,807,718, and 5,910,408). Preferred ribozymes cleave RNA or DNA substrates, and more preferably cleave RNA substrates. Ribozymes typically cleave nucleic acid substrates through recognition and binding of the target substrate with subsequent cleavage. This recognition is often based mostly  
25 on canonical or non-canonical base pair interactions. This property makes ribozymes particularly good candidates for target specific cleavage of nucleic acids because recognition of the target substrate is based on the target substrates sequence. Representative examples of how to make and use ribozymes to catalyze a variety of different reactions can be found in the following non-limiting list of United States patents: 5,646,042, 5,693,535, 5,731,295, 5,811,300,  
30 5,837,855, 5,869,253, 5,877,021, 5,877,022, 5,972,699, 5,972,704, 5,989,906, and 6,017,756.

102. Triplex forming functional nucleic acid molecules are molecules that can interact with either double-stranded or single-stranded nucleic acid. When triplex molecules interact with a target region, a structure called a triplex is formed, in which there are three strands of

DNA forming a complex dependant on both Watson-Crick and Hoogsteen base-pairing. Triplex molecules are preferred because they can bind target regions with high affinity and specificity. It is preferred that the triplex forming molecules bind the target molecule with a  $k_d$  less than  $10^{-6}$ ,  $10^{-8}$ ,  $10^{-10}$ , or  $10^{-12}$ . Representative examples of how to make and use triplex forming molecules to bind a variety of different target molecules can be found in the following non-limiting list of United States patents: 5,176,996, 5,645,985, 5,650,316, 5,683,874, 5,693,773, 5,834,185, 5,869,246, 5,874,566, and 5,962,426.

103. External guide sequences (EGSs) are molecules that bind a target nucleic acid molecule forming a complex, and this complex is recognized by RNase P, which cleaves the target molecule. EGSs can be designed to specifically target a RNA molecule of choice. RNase P aids in processing transfer RNA (tRNA) within a cell. Bacterial RNase P can be recruited to cleave virtually any RNA sequence by using an EGS that causes the target RNA:EGS complex to mimic the natural tRNA substrate. (WO 92/03566 by Yale, and Forster and Altman, Science 238:407-409 (1990)).

104. Similarly, eukaryotic EGS/RNase P-directed cleavage of RNA can be utilized to cleave desired targets within eukarotic cells. (Yuan et al., Proc. Natl. Acad. Sci. USA 89:8006-8010 (1992); WO 93/22434 by Yale; WO 95/24489 by Yale; Yuan and Altman, EMBO J 14:159-168 (1995), and Carrara et al., Proc. Natl. Acad. Sci. (USA) 92:2627-2631 (1995)). Representative examples of how to make and use EGS molecules to facilitate cleavage of a variety of different target molecules be found in the following non-limiting list of United States patents: 5,168,053, 5,624,824, 5,683,873, 5,728,521, 5,869,248, and 5,877,162.

#### ***(b) Antibodies***

105. Disclosed are monoclonal and polyclonal as well as chimeric variants of these, that bind DHR96 or variants or fragments thereof. Also disclosed are monoclonal and polyclonal antibodies that bind DHR96 or variants or fragments thereof that inhibit DHR96 activity in, for example, the xenobiotic pathways disclosed herein. Various assays are disclosed herein that can be used to identify these antibodies, such as the nutritional viability assay disclosed herein or the sensitivity to toxins assay disclosed herein.

106. As used herein, the term "antibody" encompasses, but is not limited to, whole immunoglobulin (i.e., an intact antibody) of any class. Native antibodies are usually heterotetrameric glycoproteins, composed of two identical light (L) chains and two identical heavy (H) chains. Typically, each light chain is linked to a heavy chain by one covalent disulfide bond, while the number of disulfide linkages varies between the heavy chains of different

immunoglobulin isotypes. Each heavy and light chain also has regularly spaced intrachain disulfide bridges. Each heavy chain has at one end a variable domain (V(H)) followed by a number of constant domains. Each light chain has a variable domain at one end (V(L)) and a constant domain at its other end; the constant domain of the light chain is aligned with the first constant domain of the heavy chain, and the light chain variable domain is aligned with the variable domain of the heavy chain. Particular amino acid residues are believed to form an interface between the light and heavy chain variable domains. The light chains of antibodies from any vertebrate species can be assigned to one of two clearly distinct types, called kappa (k) and lambda (l), based on the amino acid sequences of their constant domains. Depending on the amino acid sequence of the constant domain of their heavy chains, immunoglobulins can be assigned to different classes. There are five major classes of human immunoglobulins: IgA, IgD, IgE, IgG and IgM, and several of these may be further divided into subclasses (isotypes), e.g., IgG-1, IgG-2, IgG-3, and IgG-4; IgA-1 and IgA-2. One skilled in the art would recognize the comparable classes for mouse. The heavy chain constant domains that correspond to the different classes of immunoglobulins are called alpha, delta, epsilon, gamma, and mu, respectively.

107. The term "variable" is used herein to describe certain portions of the variable domains that differ in sequence among antibodies and are used in the binding and specificity of each particular antibody for its particular antigen. However, the variability is not usually evenly distributed through the variable domains of antibodies. It is typically concentrated in three segments called complementarity determining regions (CDRs) or hypervariable regions both in the light chain and the heavy chain variable domains. The more highly conserved portions of the variable domains are called the framework (FR). The variable domains of native heavy and light chains each comprise four FR regions, largely adopting a b-sheet configuration, connected by three CDRs, which form loops connecting, and in some cases forming part of, the b-sheet structure. The CDRs in each chain are held together in close proximity by the FR regions and, with the CDRs from the other chain, contribute to the formation of the antigen binding site of antibodies (see Kabat E. A. et al., "Sequences of Proteins of Immunological Interest," National Institutes of Health, Bethesda, Md. (1987)). The constant domains are not involved directly in binding an antibody to an antigen, but exhibit various effector functions, such as participation of the antibody in antibody-dependent cellular toxicity.

108. As used herein, the term "antibody or fragments thereof" encompasses chimeric antibodies and hybrid antibodies, with dual or multiple antigen or epitope specificities, and

fragments, such as F(ab')<sub>2</sub>, Fab', Fab and the like, including hybrid fragments. Thus, fragments of the antibodies that retain the ability to bind their specific antigens are provided. For example, fragments of antibodies which maintain binding activity to the DHR96 or variants or fragments thereof are included within the meaning of the term "antibody or fragment thereof." Such antibodies and fragments can be made by techniques known in the art and can be screened for specificity and activity according to the methods set forth in the Examples and in general methods for producing antibodies and screening antibodies for specificity and activity (See Harlow and Lane. Antibodies, A Laboratory Manual. Cold Spring Harbor Publications, New York, (1988)).

109. Also included within the meaning of "antibody or fragments thereof" are conjugates of antibody fragments and antigen binding proteins (single chain antibodies) as described, for example, in U.S. Pat. No. 4,704,692, the contents of which are hereby incorporated by reference.

110. Optionally, the antibodies are generated in other species and "humanized" for administration in humans. Humanized forms of non-human (e.g., murine) antibodies are chimeric immunoglobulins, immunoglobulin chains or fragments thereof (such as Fv, Fab, Fab', F(ab')<sub>2</sub>, or other antigen-binding subsequences of antibodies) which contain minimal sequence derived from non-human immunoglobulin. Humanized antibodies include human immunoglobulins (recipient antibody) in which residues from a complementary determining region (CDR) of the recipient are replaced by residues from a CDR of a non-human species (donor antibody) such as mouse, rat or rabbit having the desired specificity, affinity and capacity. In some instances, Fv framework residues of the human immunoglobulin are replaced by corresponding non-human residues. Humanized antibodies may also comprise residues that are found neither in the recipient antibody nor in the imported CDR or framework sequences. In general, the humanized antibody will comprise substantially all of at least one, and typically two, variable domains, in which all or substantially all of the CDR regions correspond to those of a non-human immunoglobulin and all or substantially all of the FR regions are those of a human immunoglobulin consensus sequence. The humanized antibody optimally also will comprise at least a portion of an immunoglobulin constant region (Fc), typically that of a human immunoglobulin (Jones et al., Nature, 321:522-525 (1986); Riechmann et al., Nature, 332:323-327 (1988); and Presta, Curr. Op. Struct. Biol., 2:593-596 (1992)).

111. Methods for humanizing non-human antibodies are well known in the art. Generally, a humanized antibody has one or more amino acid residues introduced into it from a

source that is non-human. These non-human amino acid residues are often referred to as “import” residues, which are typically taken from an “import” variable domain. Humanization can be essentially performed following the method of Winter and co-workers (Jones et al., Nature, 321:522-525 (1986); Riechmann et al., Nature, 332:323-327 (1988); Verhoeyen et al., Science, 239:1534-1536 (1988)), by substituting rodent CDRs or CDR sequences for the corresponding sequences of a human antibody. Accordingly, such “humanized” antibodies are chimeric antibodies (U.S. Pat. No. 4,816,567), wherein substantially less than an intact human variable domain has been substituted by the corresponding sequence from a non-human species. In practice, humanized antibodies are typically human antibodies in which some CDR residues and possibly some FR residues are substituted by residues from analogous sites in rodent antibodies.

112. The choice of human variable domains, both light and heavy, to be used in making the humanized antibodies is very important in order to reduce antigenicity. According to the “best-fit” method, the sequence of the variable domain of a rodent antibody is screened against the entire library of known human variable domain sequences. The human sequence which is closest to that of the rodent is then accepted as the human framework (FR) for the humanized antibody (Sims et al., J. Immunol., 151:2296 (1993) and Chothia et al., J. Mol. Biol., 196:901 (1987)). Another method uses a particular framework derived from the consensus sequence of all human antibodies of a particular subgroup of light or heavy chains. The same framework may be used for several different humanized antibodies (Carter et al., Proc. Natl. Acad. Sci. USA, 89:4285 (1992); Presta et al., J. Immunol., 151:2623 (1993)).

113. It is further important that antibodies be humanized with retention of high affinity for the antigen and other favorable biological properties. To achieve this goal, according to a preferred method, humanized antibodies are prepared by a process of analysis of the parental sequences and various conceptual humanized products using three dimensional models of the parental and humanized sequences. Three dimensional immunoglobulin models are commonly available and are familiar to those skilled in the art. Computer programs are available which illustrate and display probable three-dimensional conformational structures of selected candidate immunoglobulin sequences. Inspection of these displays permits analysis of the likely role of the residues in the functioning of the candidate immunoglobulin sequence, i.e., the analysis of residues that influence the ability of the candidate immunoglobulin to bind its antigen. In this way, FR residues can be selected and combined from the consensus and import sequence so that the desired antibody characteristic, such as increased affinity for the target antigen(s), is

achieved. In general, the CDR residues are directly and most substantially involved in influencing antigen binding (see, WO 94/04679, published 3 March 1994).

114. Transgenic animals (e.g., mice) that are capable, upon immunization, of producing a full repertoire of human antibodies in the absence of endogenous immunoglobulin production can be employed. For example, it has been described that the homozygous deletion of the antibody heavy chain joining region (J(H)) gene in chimeric and germ-line mutant mice results in complete inhibition of endogenous antibody production. Transfer of the human germ-line immunoglobulin gene array in such germ-line mutant mice will result in the production of human antibodies upon antigen challenge (see, e.g., Jakobovits et al., Proc. Natl. Acad. Sci. USA, 90:2551-255 (1993); Jakobovits et al., Nature, 362:255-258 (1993); Bruggemann et al., Year in Immuno., 7:33 (1993)). Human antibodies can also be produced in phage display libraries (Hoogenboom et al., J. Mol. Biol., 227:381 (1991); Marks et al., J. Mol. Biol., 222:581 (1991)). The techniques of Cote et al. and Boerner et al. are also available for the preparation of human monoclonal antibodies (Cole et al., Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, p. 77 (1985); Boerner et al., J. Immunol., 147(1):86-95 (1991)).

115. Disclosed are hybridoma cells that produces the monoclonal antibody. The term "monoclonal antibody" as used herein refers to an antibody obtained from a substantially homogeneous population of antibodies, i.e., the individual antibodies comprising the population are identical except for possible naturally occurring mutations that may be present in minor amounts. The monoclonal antibodies herein specifically include "chimeric" antibodies in which a portion of the heavy and/or light chain is identical with or homologous to corresponding sequences in antibodies derived from a particular species or belonging to a particular antibody class or subclass, while the remainder of the chain(s) is identical with or homologous to corresponding sequences in antibodies derived from another species or belonging to another antibody class or subclass, as well as fragments of such antibodies, so long as they exhibit the desired activity (See, U.S. Pat. No. 4,816,567 and Morrison et al., Proc. Natl. Acad. Sci. USA, 81:6851-6855 (1984)).

116. Monoclonal antibodies may be prepared using hybridoma methods, such as those described by Kohler and Milstein, Nature, 256:495 (1975) or Harlow and Lane. Antibodies, A Laboratory Manual. Cold Spring Harbor Publications, New York, (1988). In a hybridoma method, a mouse or other appropriate host animal, is typically immunized with an immunizing agent to elicit lymphocytes that produce or are capable of producing antibodies that will specifically bind to the immunizing agent. Alternatively, the lymphocytes may be immunized in

vitro. Preferably, the immunizing agent comprises DHR96 or variants or fragments thereof.

Traditionally, the generation of monoclonal antibodies has depended on the availability of purified protein or peptides for use as the immunogen. More recently DNA based

immunizations have shown promise as a way to elicit strong immune responses and generate

monoclonal antibodies. In this approach, DNA-based immunization can be used, wherein DNA encoding a portion of DHR96 or variants or fragments thereof expressed as a fusion protein with human IgG1 is injected into the host animal according to methods known in the art (e.g.,

Kilpatrick KE, et al. Gene gun delivered DNA-based immunizations mediate rapid production of murine monoclonal antibodies to the Flt-3 receptor. *Hybridoma*. 1998 Dec;17(6):569-76;

Kilpatrick KE et al. High-affinity monoclonal antibodies to PED/PEA-15 generated using 5 microg of DNA. *Hybridoma*. 2000 Aug;19(4):297-302, which are incorporated herein by referenced in full for the the methods of antibody production) and as described in the examples.

117. An alternate approach to immunizations with either purified protein or DNA is to use antigen expressed in baculovirus. The advantages to this system include ease of generation, high levels of expression, and post-translational modifications that are highly similar to those seen in mammalian systems. Use of this system involves expressing domains of antibodies to DHR96 or variants or fragments thereof as fusion proteins. The antigen is produced by inserting a gene fragment in-frame between the signal sequence and the mature protein domain of the antibodies to DHR96 or variants or fragments thereof nucleotide sequence. This results in the display of the foreign proteins on the surface of the virion. This method allows immunization with whole virus, eliminating the need for purification of target antigens.

118. Generally, either peripheral blood lymphocytes ("PBLs") are used in methods of producing monoclonal antibodies if cells of human origin are desired, or spleen cells or lymph node cells are used if non-human mammalian sources are desired. The lymphocytes are then fused with an immortalized cell line using a suitable fusing agent, such as polyethylene glycol, to form a hybridoma cell (Goding, "Monoclonal Antibodies: Principles and Practice" Academic Press, (1986) pp. 59-103). Immortalized cell lines are usually transformed mammalian cells, including myeloma cells of rodent, bovine, equine, and human origin. Usually, rat or mouse myeloma cell lines are employed. The hybridoma cells may be cultured in a suitable culture medium that preferably contains one or more substances that inhibit the growth or survival of the unfused, immortalized cells. For example, if the parental cells lack the enzyme hypoxanthine guanine phosphoribosyl transferase (HGPRT or HPRT), the culture medium for the hybridomas typically will include hypoxanthine, aminopterin, and thymidine ("HAT medium"), which

substances prevent the growth of HGPRT-deficient cells. Preferred immortalized cell lines are those that fuse efficiently, support stable high level expression of antibody by the selected antibody-producing cells, and are sensitive to a medium such as HAT medium. More preferred immortalized cell lines are murine myeloma lines, which can be obtained, for instance, from the Salk Institute Cell Distribution Center, San Diego, Calif. and the American Type Culture Collection, Rockville, Md. Human myeloma and mouse-human heteromyeloma cell lines also have been described for the production of human monoclonal antibodies (Kozbor, J. Immunol., 133:3001 (1984); Brodeur et al., "Monoclonal Antibody Production Techniques and Applications" Marcel Dekker, Inc., New York, (1987) pp. 51-63). The culture medium in which the hybridoma cells are cultured can then be assayed for the presence of monoclonal antibodies directed against DHR96 or variants or fragments thereof. Preferably, the binding specificity of monoclonal antibodies produced by the hybridoma cells is determined by immunoprecipitation or by an in vitro binding assay, such as radioimmunoassay (RIA) or enzyme-linked immunoabsorbent assay (ELISA). Such techniques and assays are known in the art, and are described further in the Examples below or in Harlow and Lane "Antibodies, A Laboratory Manual" Cold Spring Harbor Publications, New York, (1988).

119. After the desired hybridoma cells are identified, the clones may be subcloned by limiting dilution or FACS sorting procedures and grown by standard methods. Suitable culture media for this purpose include, for example, Dulbecco's Modified Eagle's Medium and RPMI-1640 medium. Alternatively, the hybridoma cells may be grown in vivo as ascites in a mammal.

120. The monoclonal antibodies secreted by the subclones may be isolated or purified from the culture medium or ascites fluid by conventional immunoglobulin purification procedures such as, for example, protein A-Sepharose, protein G, hydroxylapatite chromatography, gel electrophoresis, dialysis, or affinity chromatography.

121. The monoclonal antibodies may also be made by recombinant DNA methods, such as those described in U.S. Pat. No. 4,816,567. DNA encoding the monoclonal antibodies can be readily isolated and sequenced using conventional procedures (e.g., by using oligonucleotide probes that are capable of binding specifically to genes encoding the heavy and light chains of murine antibodies). The hybridoma cells serve as a preferred source of such DNA. Once isolated, the DNA may be placed into expression vectors, which are then transfected into host cells such as simian COS cells, Chinese hamster ovary (CHO) cells, plasmacytoma cells, or myeloma cells that do not otherwise produce immunoglobulin protein, to obtain the synthesis of monoclonal antibodies in the recombinant host cells. The DNA also may

be modified, for example, by substituting the coding sequence for human heavy and light chain constant domains in place of the homologous murine sequences (U.S. Pat. No. 4,816,567) or by covalently joining to the immunoglobulin coding sequence all or part of the coding sequence for a non-immunoglobulin polypeptide. Optionally, such a non-immunoglobulin polypeptide is substituted for the constant domains of an antibody or substituted for the variable domains of one antigen-combining site of an antibody to create a chimeric bivalent antibody comprising one antigen-combining site having specificity for DHR96 or variants or fragments thereof and another antigen-combining site having specificity for a different antigen.

122. *In vitro* methods are also suitable for preparing monovalent antibodies. Digestion of antibodies to produce fragments thereof, particularly, Fab fragments, can be accomplished using routine techniques known in the art. For instance, digestion can be performed using papain. Examples of papain digestion are described in WO 94/29348 published Dec. 22, 1994, U.S. Pat. No. 4,342,566, and Harlow and Lane, *Antibodies, A Laboratory Manual*, Cold Spring Harbor Publications, New York, (1988). Papain digestion of antibodies typically produces two identical antigen binding fragments, called Fab fragments, each with a single antigen binding site, and a residual Fc fragment. Pepsin treatment yields a fragment, called the F(ab')<sub>2</sub> fragment, that has two antigen combining sites and is still capable of cross-linking antigen.

123. The Fab fragments produced in the antibody digestion also contain the constant domains of the light chain and the first constant domain of the heavy chain. Fab' fragments differ from Fab fragments by the addition of a few residues at the carboxy terminus of the heavy chain domain including one or more cysteines from the antibody hinge region. The F(ab')<sub>2</sub> fragment is a bivalent fragment comprising two Fab' fragments linked by a disulfide bridge at the hinge region. Fab'-SH is the designation herein for Fab' in which the cysteine residue(s) of the constant domains bear a free thiol group. Antibody fragments originally were produced as pairs of Fab' fragments which have hinge cysteines between them. Other chemical couplings of antibody fragments are also known.

124. An isolated immunogenically specific paratope or fragment of the antibody is also provided. A specific immunogenic epitope of the antibody can be isolated from the whole antibody by chemical or mechanical disruption of the molecule. The purified fragments thus obtained are tested to determine their immunogenicity and specificity by the methods taught herein. Immunoreactive paratopes of the antibody, optionally, are synthesized directly. An immunoreactive fragment is defined as an amino acid sequence of at least about two to five consecutive amino acids derived from the antibody amino acid sequence.

125. One method of producing proteins comprising the antibodies is to link two or more peptides or polypeptides together by protein chemistry techniques. For example, peptides or polypeptides can be chemically synthesized using currently available laboratory equipment using either Fmoc (9-fluorenylmethyloxycarbonyl) or Boc (tert -butyloxycarbonyl) chemistry. (Applied Biosystems, Inc., Foster City, CA). One skilled in the art can readily appreciate that a peptide or polypeptide corresponding to the antibody, for example, can be synthesized by standard chemical reactions. For example, a peptide or polypeptide can be synthesized and not cleaved from its synthesis resin whereas the other fragment of an antibody can be synthesized and subsequently cleaved from the resin, thereby exposing a terminal group which is functionally blocked on the other fragment. By peptide condensation reactions, these two fragments can be covalently joined via a peptide bond at their carboxyl and amino termini, respectively, to form an antibody, or fragment thereof. (Grant GA (1992) *Synthetic Peptides: A User Guide*. W.H. Freeman and Co., N.Y. (1992); Bodansky M and Trost B., Ed. (1993) *Principles of Peptide Synthesis*. Springer-Verlag Inc., NY. Alternatively, the peptide or polypeptide is independently synthesized in vivo as described above. Once isolated, these independent peptides or polypeptides may be linked to form an antibody or fragment thereof via similar peptide condensation reactions.

126. For example, enzymatic ligation of cloned or synthetic peptide segments allow relatively short peptide fragments to be joined to produce larger peptide fragments, polypeptides or whole protein domains (Abrahmsen L et al., *Biochemistry*, 30:4151 (1991)). Alternatively, native chemical ligation of synthetic peptides can be utilized to synthetically construct large peptides or polypeptides from shorter peptide fragments. This method consists of a two step chemical reaction (Dawson et al. *Synthesis of Proteins by Native Chemical Ligation*. *Science*, 266:776-779 (1994)). The first step is the chemoselective reaction of an unprotected synthetic peptide-alpha-thioester with another unprotected peptide segment containing an amino-terminal Cys residue to give a thioester-linked intermediate as the initial covalent product. Without a change in the reaction conditions, this intermediate undergoes spontaneous, rapid intramolecular reaction to form a native peptide bond at the ligation site. Application of this native chemical ligation method to the total synthesis of a protein molecule is illustrated by the preparation of human interleukin 8 (IL-8) (Baggiolini M et al. (1992) *FEBS Lett.* 307:97-101; Clark-Lewis I et al., *J.Biol.Chem.*, 269:16075 (1994); Clark-Lewis I et al., *Biochemistry*, 30:3128 (1991); Rajarathnam K et al., *Biochemistry* 33:6623-30 (1994)).

127. Alternatively, unprotected peptide segments are chemically linked where the bond formed between the peptide segments as a result of the chemical ligation is an unnatural (non-peptide) bond (Schnolzer, M et al. Science, 256:221 (1992)). This technique has been used to synthesize analogs of protein domains as well as large amounts of relatively pure proteins with full biological activity (deLisle Milton RC et al., Techniques in Protein Chemistry IV. Academic Press, New York, pp. 257-267 (1992)).

128. Also disclosed are fragments of antibodies which have bioactivity. The polypeptide fragments can be recombinant proteins obtained by cloning nucleic acids encoding the polypeptide in an expression system capable of producing the polypeptide fragments thereof, such as an adenovirus or baculovirus expression system. For example, one can determine the active domain of an antibody from a specific hybridoma that can cause a biological effect associated with the interaction of the antibody with DHR96 or variants or fragments thereof. For example, amino acids found to not contribute to either the activity or the binding specificity or affinity of the antibody can be deleted without a loss in the respective activity. For example, in various embodiments, amino or carboxy-terminal amino acids are sequentially removed from either the native or the modified non-immunoglobulin molecule or the immunoglobulin molecule and the respective activity assayed in one of many available assays. In another example, a fragment of an antibody comprises a modified antibody wherein at least one amino acid has been substituted for the naturally occurring amino acid at a specific position, and a portion of either amino terminal or carboxy terminal amino acids, or even an internal region of the antibody, has been replaced with a polypeptide fragment or other moiety, such as biotin, which can facilitate in the purification of the modified antibody. For example, a modified antibody can be fused to a maltose binding protein, through either peptide chemistry or cloning the respective nucleic acids encoding the two polypeptide fragments into an expression vector such that the expression of the coding region results in a hybrid polypeptide. The hybrid polypeptide can be affinity purified by passing it over an amylose affinity column, and the modified antibody receptor can then be separated from the maltose binding region by cleaving the hybrid polypeptide with the specific protease factor Xa. (See, for example, New England Biolabs Product Catalog, 1996, pg. 164.). Similar purification procedures are available for isolating hybrid proteins from eukaryotic cells as well.

129. The fragments, whether attached to other sequences or not, include insertions, deletions, substitutions, or other selected modifications of particular regions or specific amino acids residues, provided the activity of the fragment is not significantly altered or impaired

compared to the nonmodified antibody or antibody fragment. These modifications can provide for some additional property, such as to remove or add amino acids capable of disulfide bonding, to increase its bio-longevity, to alter its secretory characteristics, etc. In any case, the fragment must possess a bioactive property, such as binding activity, regulation of binding at the binding domain, etc. Functional or active regions of the antibody may be identified by mutagenesis of a specific region of the protein, followed by expression and testing of the expressed polypeptide. Such methods are readily apparent to a skilled practitioner in the art and can include site-specific mutagenesis of the nucleic acid encoding the antigen. (Zoller MJ et al. Nucl. Acids Res. 10:6487-500 (1982).

130. A variety of immunoassay formats may be used to select antibodies that selectively bind with a particular protein, variant, or fragment. For example, solid-phase ELISA immunoassays are routinely used to select antibodies selectively immunoreactive with a protein, protein variant, or fragment thereof. See Harlow and Lane. Antibodies, A Laboratory Manual. Cold Spring Harbor Publications, New York, (1988), for a description of immunoassay formats and conditions that could be used to determine selective binding. The binding affinity of a monoclonal antibody can, for example, be determined by the Scatchard analysis of Munson et al., Anal. Biochem., 107:220 (1980).

131. Also provided is an antibody reagent kit comprising containers of the monoclonal antibody or fragment thereof and one or more reagents for detecting binding of the antibody or fragment thereof to DHR96 or variants or fragments thereof. The reagents can include, for example, fluorescent tags, enzymatic tags, or other tags. The reagents can also include secondary or tertiary antibodies or reagents for enzymatic reactions, wherein the enzymatic reactions produce a product that can be visualized.

***(c) Compositions identified by screening with disclosed compositions / combinatorial chemistry***

***(i) Combinatorial chemistry***

132. The disclosed compositions can be used as targets for any combinatorial technique to identify molecules or macromolecular molecules that interact with the disclosed compositions in a desired way. The nucleic acids, peptides, and related molecules disclosed herein, such as DHR96 or variants or fragments thereof, can be used as targets for the combinatorial approaches. Also disclosed are the compositions that are identified through combinatorial techniques or screening techniques in which the compositions, such as DHR96 or

variants or fragments thereof, or portions thereof, are used as the target in a combinatorial or screening protocol.

133. It is understood that when using the disclosed compositions in combinatorial techniques or screening methods, molecules, such as macromolecular molecules, will be identified that have particular desired properties such as inhibition or stimulation or the target molecule's function. The molecules identified and isolated when using the disclosed compositions, such as, DHR96 or variants or fragments thereof, are also disclosed. Thus, the products produced using the combinatorial or screening approaches that involve the disclosed compositions, such as, DHR96 or variants or fragments thereof, are also considered herein disclosed.

134. It is understood that the disclosed methods for identifying molecules that inhibit the interactions between, for example, DHR96 or variants or fragments thereof, can be performed using high through put means. For example, putative inhibitors can be identified using Fluorescence Resonance Energy Transfer (FRET) to quickly identify interactions. The underlying theory of the techniques is that when two molecules are close in space, ie, interacting at a level beyond background, a signal is produced or a signal can be quenched. Then, a variety of experiments can be performed, including, for example, adding in a putative inhibitor. If the inhibitor competes with the interaction between the two signaling molecules, the signals will be removed from each other in space, and this will cause a decrease or an increase in the signal, depending on the type of signal used. This decrease or increasing signal can be correlated to the presence or absence of the putative inhibitor. Any signaling means can be used. For example, disclosed are methods of identifying an inhibitor of the interaction between any two of the disclosed molecules comprising, contacting a first molecule and a second molecule together in the presence of a putative inhibitor, wherein the first molecule or second molecule comprises a fluorescence donor, wherein the first or second molecule, typically the molecule not comprising the donor, comprises a fluorescence acceptor; and measuring Fluorescence Resonance Energy Transfer (FRET), in the presence of the putative inhibitor and the in absence of the putative inhibitor, wherein a decrease in FRET in the presence of the putative inhibitor as compared to FRET measurement in its absence indicates the putative inhibitor inhibits binding between the two molecules. This type of method can be performed with a cell system as well.

135. Combinatorial chemistry includes but is not limited to all methods for isolating small molecules or macromolecules that are capable of binding either a small molecule or another macromolecule, typically in an iterative process. Proteins, oligonucleotides, and sugars

are examples of macromolecules. For example, oligonucleotide molecules with a given function, catalytic or ligand-binding, can be isolated from a complex mixture of random oligonucleotides in what has been referred to as "in vitro genetics" (Szostak, TIBS 19:89, 1992).

One synthesizes a large pool of molecules bearing random and defined sequences and subjects that complex mixture, for example, approximately  $10^{15}$  individual sequences in 100  $\mu$ g of a 100 nucleotide RNA, to some selection and enrichment process. Through repeated cycles of affinity chromatography and PCR amplification of the molecules bound to the ligand on the column, Ellington and Szostak (1990) estimated that 1 in  $10^{10}$  RNA molecules folded in such a way as to bind a small molecule dyes. DNA molecules with such ligand-binding behavior have been isolated as well (Ellington and Szostak, 1992; Bock et al, 1992). Techniques aimed at similar goals exist for small organic molecules, proteins, antibodies and other macromolecules known to those of skill in the art. Screening sets of molecules for a desired activity whether based on small organic libraries, oligonucleotides, or antibodies is broadly referred to as combinatorial chemistry. Combinatorial techniques are particularly suited for defining binding interactions between molecules and for isolating molecules that have a specific binding activity, often called aptamers when the macromolecules are nucleic acids.

136. There are a number of methods for isolating proteins which either have de novo activity or a modified activity. For example, phage display libraries have been used to isolate numerous peptides that interact with a specific target. (See for example, United States Patent No. 6,031,071; 5,824,520; 5,596,079; and 5,565,332 which are herein incorporated by reference at least for their material related to phage display and methods relate to combinatorial chemistry)

137. A preferred method for isolating proteins that have a given function is described by Roberts and Szostak (Roberts R.W. and Szostak J.W. Proc. Natl. Acad. Sci. USA, 94(23)12997-302 (1997). This combinatorial chemistry method couples the functional power of proteins and the genetic power of nucleic acids. An RNA molecule is generated in which a puromycin molecule is covalently attached to the 3'-end of the RNA molecule. An *in vitro* translation of this modified RNA molecule causes the correct protein, encoded by the RNA to be translated. In addition, because of the attachment of the puromycin, a peptidyl acceptor which cannot be extended, the growing peptide chain is attached to the puromycin which is attached to the RNA. Thus, the protein molecule is attached to the genetic material that encodes it. Normal *in vitro* selection procedures can now be done to isolate functional peptides. Once the selection procedure for peptide function is complete traditional nucleic acid manipulation procedures are performed to amplify the nucleic acid that codes for the selected functional peptides. After

amplification of the genetic material, new RNA is transcribed with puromycin at the 3'-end, new peptide is translated and another functional round of selection is performed. Thus, protein selection can be performed in an iterative manner just like nucleic acid selection techniques. The peptide which is translated is controlled by the sequence of the RNA attached to the puromycin.

5 This sequence can be anything from a random sequence engineered for optimum translation (i.e. no stop codons etc.) or it can be a degenerate sequence of a known RNA molecule to look for improved or altered function of a known peptide. The conditions for nucleic acid amplification and in vitro translation are well known to those of ordinary skill in the art and are preferably performed as in Roberts and Szostak (Roberts R.W. and Szostak J.W. Proc. Natl. Acad. Sci.  
10 USA, 94(23)12997-302 (1997)).

138. Another preferred method for combinatorial methods designed to isolate peptides is described in Cohen et al. (Cohen B.A., et al., Proc. Natl. Acad. Sci. USA 95(24):14272-7 (1998)). This method utilizes and modifies two-hybrid technology. Yeast two-hybrid systems are useful for the detection and analysis of protein:protein interactions. The two-hybrid system,  
15 initially described in the yeast *Saccharomyces cerevisiae*, is a powerful molecular genetic technique for identifying new regulatory molecules, specific to the protein of interest (Fields and Song, *Nature* 340:245-6 (1989)). Cohen et al., modified this technology so that novel interactions between synthetic or engineered peptide sequences could be identified which bind a molecule of choice. The benefit of this type of technology is that the selection is done in an  
20 intracellular environment. The method utilizes a library of peptide molecules that attached to an acidic activation domain. A peptide of choice, for example, of DHR96 or variants or fragments thereof, is attached to a DNA binding domain of a transcriptional activation protein, such as Gal 4. By performing the two-hybrid technique on this type of system, molecules that bind DHR96 or variants or fragments thereof can be identified.

25 139. Using methodology well known to those of skill in the art, in combination with various combinatorial libraries, one can isolate and characterize those small molecules or macromolecules, which bind to or interact with the desired target. The relative binding affinity of these compounds can be compared and optimum compounds identified using competitive binding studies, which are well known to those of skill in the art.

30 140. Techniques for making combinatorial libraries and screening combinatorial libraries to isolate molecules which bind a desired target are well known to those of skill in the art. Representative techniques and methods can be found in but are not limited to United States patents 5,084,824, 5,288,514, 5,449,754, 5,506,337, 5,539,083, 5,545,568, 5,556,762, 5,565,324,

5,565,332, 5,573,905, 5,618,825, 5,619,680, 5,627,210, 5,646,285, 5,663,046, 5,670,326, 5,677,195, 5,683,899, 5,688,696, 5,688,997, 5,698,685, 5,712,146, 5,721,099, 5,723,598, 5,741,713, 5,792,431, 5,807,683, 5,807,754, 5,821,130, 5,831,014, 5,834,195, 5,834,318, 5,834,588, 5,840,500, 5,847,150, 5,856,107, 5,856,496, 5,859,190, 5,864,010, 5,874,443, 5,877,214, 5,880,972, 5,886,126, 5,886,127, 5,891,737, 5,916,899, 5,919,955, 5,925,527, 5,939,268, 5,942,387, 5,945,070, 5,948,696, 5,958,702, 5,958,792, 5,962,337, 5,965,719, 5,972,719, 5,976,894, 5,980,704, 5,985,356, 5,999,086, 6,001,579, 6,004,617, 6,008,321, 6,017,768, 6,025,371, 6,030,917, 6,040,193, 6,045,671, 6,045,755, 6,060,596, and 6,061,636.

141. Combinatorial libraries can be made from a wide array of molecules using a number of different synthetic techniques. For example, libraries containing fused 2,4-pyrimidinediones (United States patent 6,025,371) dihydrobenzopyrans (United States Patent 6,017,768 and 5,821,130), amide alcohols (United States Patent 5,976,894), hydroxy-amino acid amides (United States Patent 5,972,719) carbohydrates (United States patent 5,965,719), 1,4-benzodiazepin-2,5-diones (United States patent 5,962,337), cyclics (United States patent 5,958,792), biaryl amino acid amides (United States patent 5,948,696), thiophenes (United States patent 5,942,387), tricyclic Tetrahydroquinolines (United States patent 5,925,527), benzofurans (United States patent 5,919,955), isoquinolines (United States patent 5,916,899), hydantoin and thiohydantoin (United States patent 5,859,190), indoles (United States patent 5,856,496), imidazol-pyrido-indole and imidazol-pyrido-benzothiophenes (United States patent 5,856,107) substituted 2-methylene-2, 3-dihydrothiazoles (United States patent 5,847,150), quinolines (United States patent 5,840,500), PNA (United States patent 5,831,014), containing tags (United States patent 5,721,099), polyketides (United States patent 5,712,146), morpholino-subunits (United States patent 5,698,685 and 5,506,337), sulfamides (United States patent 5,618,825), and benzodiazepines (United States patent 5,288,514).

142. As used herein combinatorial methods and libraries included traditional screening methods and libraries as well as methods and libraries used in iterative processes.

*(ii) Computer assisted drug design*

143. The disclosed compositions can be used as targets for any molecular modeling technique to identify either the structure of the disclosed compositions or to identify potential or actual molecules, such as small molecules, which interact in a desired way with the disclosed compositions. The nucleic acids, peptides, and related molecules disclosed herein, such as DHR96 or variants or fragments thereof, can be used as targets in any molecular modeling program or approach.

144. It is understood that when using the disclosed compositions in modeling techniques, molecules, such as macromolecular molecules, will be identified that have particular desired properties such as inhibition or stimulation or the target molecule's function. The molecules identified and isolated when using the disclosed compositions, such as, DHR96 or  
5 variants or fragments thereof, are also disclosed. Thus, the products produced using the molecular modeling approaches that involve the disclosed compositions, such as, DHR96 or variants or fragments thereof, are also considered herein disclosed.

145. Thus, one way to isolate molecules that bind a molecule of choice is through rational design. This is achieved through structural information and computer modeling.

10 Computer modeling technology allows visualization of the three-dimensional atomic structure of a selected molecule and the rational design of new compounds that will interact with the molecule. The three-dimensional construct typically depends on data from x-ray crystallographic analyses or NMR imaging of the selected molecule. The molecular dynamics require force field data. The computer graphics systems enable prediction of how a new compound will link to the  
15 target molecule and allow experimental manipulation of the structures of the compound and target molecule to perfect binding specificity. Prediction of what the molecule-compound interaction will be when small changes are made in one or both requires molecular mechanics software and computationally intensive computers, usually coupled with user-friendly, menu-driven interfaces between the molecular design program and the user.

20 146. Examples of molecular modeling systems are the CHARMM and QUANTA programs, Polygen Corporation, Waltham, MA. CHARMM performs the energy minimization and molecular dynamics functions. QUANTA performs the construction, graphic modeling and analysis of molecular structure. QUANTA allows interactive construction, modification, visualization, and analysis of the behavior of molecules with each other.

25 147. A number of articles review computer modeling of drugs interactive with specific proteins, such as Rotivinen, et al., 1988 *Acta Pharmaceutica Fennica* 97, 159-166; Ripka, *New Scientist* 54-57 (June 16, 1988); McKinaly and Rossmann, 1989 *Annu. Rev. Pharmacol. Toxicol.* 29, 111-122; Perry and Davies, QSAR: Quantitative Structure-Activity Relationships in Drug Design pp. 189-193 (Alan R. Liss, Inc. 1989); Lewis and Dean, 1989 *Proc. R. Soc. Lond.* 236, 125-140 and 141-162; and, with respect to a model enzyme for nucleic acid  
30 components, Askew, et al., 1989 *J. Am. Chem. Soc.* 111, 1082-1090. Other computer programs that screen and graphically depict chemicals are available from companies such as BioDesign, Inc., Pasadena, CA., Allelix, Inc, Mississauga, Ontario, Canada, and Hypercube, Inc.,

Cambridge, Ontario. Although these are primarily designed for application to drugs specific to particular proteins, they can be adapted to design of molecules specifically interacting with specific regions of DNA or RNA, once that region is identified.

148. Although described above with reference to design and generation of compounds which could alter binding, one could also screen libraries of known compounds, including natural products or synthetic chemicals, and biologically active materials, including proteins, for compounds which alter substrate binding or enzymatic activity.

#### **(5) Insects that can be targeted**

149. Arthropods include Crustacea, which are things like prawns, crabs and woodlice; Myriapoda, which are centipedes, millipedes and such; Chelicerata (Arachnida), which are spiders, scorpions and harvestmen etc., and Uniramia (Insecta), which are things like beetles, bees and flies.

150. Insects are found in the phylum Arthropoda, Subphylum Insecta (also often called a class), Class Hexapoda, and Subclasses Apterygota, Exopterygota, and Endopterygota. The Apterygota includes the orders Protura, Collembola (Springtails), Thysanura (Silverfish), Diplura (Two Pronged Bristle-tails). The Exopterygota includes the orders Ephemeroptera (Mayflies), Odonata (Dragonflies), Plecoptera (Stoneflies), Grylloblatodea, Orthoptera, Phasmida (Stick-Insects), Dermaptera (Earwigs), Embioptera (Web Spinners), Dictyoptera (Cockroaches and Mantids), Isoptera (Termites), Zoraptera, Psocoptera (Bark and Book Lice), Mallophaga (Biting Lice), Siphunculata (Sucking Lice), Hemiptera (True Bugs) Thysanoptera, The Endopterygota includes the orders Neuropter (Lacewings), Coleoptera (Beetles), Strepsiptera (Stylops), Mecoptera (Scorpionflies), Siphonaptera (Fleas), Diptera (True Flies which are unusual in that they only have one pair of functional wings. The other pair is reduced to a pair of knoblike organs, called halteres, which play a part in stabilizing these insects during flight. True flies include house flies and bluebottles, mosquitoes, horseflies, midges, and antler-headed flies), Lepidoptera (Butterflies and Moths), Trichoptera (Caddis Flies), and Hymenoptera (Ants Bees and Wasps).

#### **(6) Exemplary pesticides that can be used in combination**

151. The disclosed compositions, such as DHR96 inhibitors can be combined with any pesticide or class of pesticides. For example, the DHR96 inhibitors can be combined with a pesticide that invokes the xenobiotic pathway. The DHR96 inhibitors can also be combined with any pesticide that effects the expression of a gene in the following four families, cytochrome P450s, carboxylesterases, glutathione S-transferases, and UDP-glucuronosyltransferases When it

is unknown which xenobiotic genes are affected by the pesticide, this can be determined by observing whether the pesticide turns on one or more genes that are in the xenobiotic pathway, by for example, microarray technology, or any other technology that determines gene expression, such as RT-PCR. In certain embodiments, when a particular gene product is specifically  
5 overexpressed in a resistant line of insects, that gene product can be considered a xenobiotic gene. Other examples, such as cuticle proteins and a serum carrier protein, were seen in the microarray experiments as well. In other embodiments any encoded protein that confers resistance to a toxic compound can be considered a xenobiotic compound.

152. There are many different pesticides that are relatively common chemicals, such as  
10 arsenicals, petroleum oils, nicotine, pyrethrum, rotenone, sulfur, hydrogen cyanide gas, and cryolite. However, most pesticides are non-natural chemically synthesized compounds. For example, there are different classes and subclasses of pesticides, such as organochlorines, examples of which are diphenyl aliphatics, hexachlorocyclohexane (HCH) or benzenehexachloride (BHC), Cyclodienes, Polychloroterpenes, organophosphates (OPs)  
15 examples of which are esters of phosphorus, organosulfers, carbamates, formamidines, dinitrophenols, organotin, pyrethroids, nicotinoids (also known as nitro-quanidines, neonicotinyls, neonicotinoids, chloronicotines, or chloronicotinyls), spinosyns, fiproles (or Phenylpyrazoles), pyrroles, pyrazoles, pyridazinones, quinazolines, benzoylureas, botanicals, (natural insecticides), synergists or activators, antibiotics, fumigants, insect repellants, and  
20 inorganics.

153. Another way of classifying insecticides is by their mode of action, for example, sodium and/or potassium channel inhibitors, buerotoxins, GABA (gamma-aminobutyric acid) receptor modulators, such as inhibitors and activators, cholinesterase (ChE) inhibitors, aliesterase inhibitors, monoamine oxidase inhibitors, oxidative phosphorylation couplers or  
25 uncouplers, adenosine triphosphate (ATP) formation inhibitors, dinitrophenol uncoupling inhibitors, axionic poisons, inhibition of postsynaptic nicotinic acetylcholine receptors, inhibiting of binding of acetylcholine in nicotinic acetylcholine receptors at the postsynaptic cell, inhibition of gamma-aminobutyric acid- (GABA) regulated chloride channels in neurons, inhibitors of mitochondrial electron transport at the NADH-CoQ reductase site, general  
30 inhibitors of mitochondrial electron transport at Site 1, insect growth regulators (IGR, inhibitors of various life cycles and stages in the insect), chitin synthesis inhibitors, inhibitors of exoskeleton development, respiratory enzyme inhibitors, inhibitors of the interaction between NAD<sup>+</sup> and coenzyme Q, inhibitors of molting, inhibitors of the biosynthesis or metabolism of

ecdysone, synergists, such as inhibitors of cytochrome P-450 dependent polysubstrate monooxygenases (PSMOs), and narcotics, calcium channel inhibitors, and repellants.

154. Examples of organochlorines are (chlorinated hydrocarbons, chlorinated organics, chlorinated insecticides, and chlorinated synthetics) Diphenyl Aliphatics, such as DDT, DDD, dicofol, ethylan, chlorobenzilate, and methoxychlor, Hexchlorocyclohexanes (HCH) or benzenehexachloride (BHC), which are typically gamma isomers, such as lindane, Cyclodienes, such as chlordane, aldrin and dieldrin, heptachlor, endrin, mirex, endosulfan, and chlordecone (Kepone®), and Polychloroterpenes, such as toxaphene and strobane.

155. Examples of organophosphates (OPs) examples of which are esters of phosphorus, (also called organic phosphates, phosphorus insecticides, nerve gas relatives, and phosphoric acid esters) derived from phosphorus acids, such as sarin, soman, and tabun, subclasses included phosphates, phospho-nates, phosphorothioates, phosphorodithioates, phosphorothiolates and phosphoramidates. There are also aliphatic, phenyl, and heterocyclic derivatives. The aliphatics include TEPP, malathion, trichlorfon (Dylox®), monocrotophos (Azodrin®), dimethoate (Cygon®), oxydemetonmethyl (Meta Systox®), dimethoate (Cygon®), dicrotophos (Bidrin®), disulfoton (Di-Syston®), dichlorvos (Vapona®), mevinphos (Phosdrin®), methamidophos (Monitor®), and acephate (Orthene®). The Phenyl derivatives parathion (ethyl parathion), methyl parathion, profenofos (Curacron®), sulprofos (Bolstar®), isofenphos (Oftanol®, Pryfon®), fenitrothion (Sumithion®), fenthion (Dasanit®), famphur (Cyflee® and Warbex®). The Heterocyclic derivatives include diazinon, azinphos-methyl (Guthion®), azinphos-ethyl (Acifon®, Gusathion®), chlorpyrifos (Dursban®, Lorsban®, Lock-On®), methidathion (Supracide®), phosmet (Imidan®), isazophos (Brace®, Triumph®), and chlorpyrifos-methyl (Reldan®).

156. Examples of organosulfers typically contain two phenyl rings, resembling DDT, with sulfur in place of carbon as the central atom, and include tetradifon (Tedion®), propargite (Omite®, Comite®), and ovex (Ovotran®).

157. Examples of carbamates are derivatives of carbamic acid and include carbaryl (Sevin®), methomyl (Lannate®), carbofuran (Furadan®), aldicarb (Temik®), oxamyl (Vydate®), thiodicarb (Larvin®), methiocarb (Mesurol®), propoxur (Baygon®), bendiocarb (Ficam®), carbosulfan (Advantage®), aldoxycarb (Standak®), promecarb (Carbamult®), and fenoxycarb (Logic®, Torus®).

158. Examples of formamidines include chlordimeform (Galecron®, Fundal®), formetanate (Carzol®), and amitraz (Mitac®, Ovasyn®).

159. Examples of dinitrophenols include binapacryl (Morocide®) and dinocap (Karathane®).

160. Examples of oganotins include cyhexatin (Plictran®) and Fenbutatin-oxide (Vendex®).

5 161. Examples of pyrethroids natural pyrethrum and synthetic pyrethroids including allethrin (Pynamin®), tetramethrin (Neo-Pynamin®) (1965), resmethrin (Synthrin®), bioresmethrin, Bioallethrin®, phonothrin (Sumithrin®), fenvalerate (Pydrin®, Tribute®, & Bellmark®), permethrin (Ambush®, Astro®, Dragnet®, Flee®, Pounce®, Prelude®, Talcord® & Torpedo®), bifenthrin (Capture®, Talstar®), *lambda*-cyhalothrin (Demand®, Karate®,  
10 Scimitar® & Warrior®), cypermethrin (Ammo®, Barricade®, Cymbush®, Cynoff® & Ripcord®), cyfluthrin (Baythroid®, Countdown®, Cylense®, Laser® & Tempo®), deltamethrin (Decis®) esfenvalerate (Asana®, Hallmark®), fenpropathrin (Danitol®), flucythrinate (Cybolt®, Payoff®), fluvalinate (Mavrik®, Spur®), prallethrin (Etoc®), *tau*-fluvalinate (Mavrik®) tefluthrin (Evict®, Fireban®, Force® & Raze®), tralomethrin (Scout X-TRA®, Tralex®), and  
15 *zeta*-cypermethrin (Mustang® Fury®), acrinathrin (Rufast®), and imiprothrin (Pralle®).

162. Examples of nicotinoids (also known as nitro-quanidines, neonicotinyls, neonicotinoids, chloronicotines, or chloronicotinyls) including Imidacloprid (Admire®, Confidor®, Gaucho®, Merit®, Premier®, Premise® and Provado®), acetamiprid (Mospilan®), thiamethoxam (Actara®, Platinum®), and nitenpyram (Bestguard®).

20 163. Examples of spinosyns include (Success®, Tracer Naturalyte®).

164. Examples of fiproles (or Phenylpyrazoles) include Fipronil ((Regent®, Icon®, Frontline®).

165. Examples of pyrroles include Chlorfenapyr ((Alert®, Pirate®.

25 166. Examples of pyrazoles include tebufenpyrad (Pyranica®, Masai®) and fenpyroximate (Acaban®, Dynamite®).

167. Examples of pyridazinones include Pyridaben ((Nexter®, Sanmite®).

168. Examples of quinazolines fenazaquin ((Matador®).

169. Examples of benzoylureas include triflumuron (Alsystin®), chlorfluazuron (Atabron®, Helix®), followed by teflubenzuron (Nomolt®, Dart®), hexaflumuron (Trueno®, Consult®), flufenoxuron (Cascade®), flucycloxuron (Andalin®), flurazuron, novaluron, diafenthiuron, Lufenuron (Axor®), and diflubenzuron ((Dimilin®, Adept®, Micromite®).

170. Examples of botanicals, (natural insecticides) include sulfur, tobacco, pyrethrum, derris, hellebore, quassia, camphor, and turpentine, and Pyrethrum, alkaloids, such as nicotine,

caffeine (coffee, tea), quinine (cinchona bark), morphine (opium poppy), cocaine (coca leaves), ricinine (a poison in castor oil beans), strychnine (*Strychnos nux vomica*), coniine (spotted hemlock, the poison used by Socrates), and LSD (a hallucigen from the ergot fungus attacking grain), rotenone, Limonene or d-Limonene, neem, Azadirachtin (Azatin® is marketed as an insect growth regulator, and Align® and Nemix®).

171. Examples of synergists or activators are not insecticides per se, but rather enhance the activity of insecticides having a primary insecticidal effect. Examples include, piperonyl butoxide, and contain the methylenedioxyphenyl moiety (found in sesame seed oil (*sesamin*)).

172. Examples of antibiotics include avermectins, Abamectin, Clinch®, Enamectin benzoate (Proclaim®, Denim®).

173. Examples of fumigants typically contain one or more halogens, such as methyl bromide (Aspelin and Grube 1998), ethylene dichloride, hydrogen cyanide, sulfuryl fluoride (Vikane®), Vapam®, Telone® II, D-D®, chloroethene, ethylene oxide, naphthalene crystals, paradichlorobenzene crystals, Phosphine gas (PH<sub>3</sub>) produced by aluminum or magnesium phosphide pellets.

174. Examples of insect repellants include dimethyl phthalate, Indalone®, Rutgers 612®, dibutyl phthalate, various MGK® repellents, benzyl benzoate, the military clothing repellent (N-butyl acetanilide), dimethyl carbate (Dimelone®) and diethyl toluamide (DEET, Delphene®).

175. Examples of inorganics include sulfur, mercury, boron, thallium, arsenic, antimony, selenium, and fluoride, arsenicals, including copper arsenate, Paris green, lead arsenate, and calcium arsenate, inorganic fluorides such as sodium fluoride, barium fluosilicate, sodium silicofluoride, and cryolite (Kryocide®), Boric acid, Sodium borate (disodium octaborate tetrahydrate) (Tim-Bor®, Bora-Care®), silica gels or silica aerogels, such as Dri-Die®, Drianone®, and Silikil Microcel®.

176. Other compounds not easily categorized include cyromazine (Larvadex®, Trigard®), a triazine, pyriproxyfen (Knack®, Esteem®, Archer®), insect growth inhibitors such as buprofezin (Applaud®) and thiadiazines, tetrazines, such as clofentezine (Apollo®, Acaristop®), Enzone®, sodium tetrathiocarbonate, and Clandosan®.

177. Also used are Veratrum Alkaloids, such as sabadilla, veratridine, and cevadine.

178. Also used are ryanoids, such as ryanodine, 10-(*O*-methyl)-ryanodine, 9,21-dehydroryanodine, ryanodol, and 9,21-dehydroryanodine.

179. Also used are octopamines mimics, such as amitraz® and chlordimeform.

180. Also included are respiration inhibitors, such as fenazaquin, pyridaben, amidinohydrazone, hydramethylnon and the perfluorooctanesulfonamide, and sulfluramid.

181. Also included are juvenile hormone mimics, such a juvenile hormone III, methoprene, and fenoxycarb.

182. Also included are toxins produced by *Bacillus thuringiensis*, such as Dipel®, Javelin®, Agree®.

### C. Compositions

183. Disclosed are the components to be used to prepare the disclosed compositions as well as the compositions themselves to be used within the methods disclosed herein. These and other materials are disclosed herein, and it is understood that when combinations, subsets, interactions, groups, etc. of these materials are disclosed that while specific reference of each various individual and collective combinations and permutation of these compounds may not be explicitly disclosed, each is specifically contemplated and described herein. For example, if a particular DHR96 or variants or fragments thereof is disclosed and discussed and a number of modifications that can be made to a number of molecules including the DHR96 or variants or fragments thereof are discussed, specifically contemplated is each and every combination and permutation of DHR96 or variants or fragments thereof and the modifications that are possible unless specifically indicated to the contrary. Thus, if a class of molecules A, B, and C are disclosed as well as a class of molecules D, E, and F and an example of a combination molecule, A-D is disclosed, then even if each is not individually recited each is individually and collectively contemplated meaning combinations, A-E, A-F, B-D, B-E, B-F, C-D, C-E, and C-F are considered disclosed. Likewise, any subset or combination of these is also disclosed. Thus, for example, the sub-group of A-E, B-F, and C-E would be considered disclosed. This concept applies to all aspects of this application including, but not limited to, steps in methods of making and using the disclosed compositions. Thus, if there are a variety of additional steps that can be performed it is understood that each of these additional steps can be performed with any specific embodiment or combination of embodiments of the disclosed methods.

#### 1. Sequence similarities

184. It is understood that as discussed herein the use of the terms homology and identity mean the same thing as similarity. Thus, for example, if the use of the word homology is used between two non-natural sequences it is understood that this is not necessarily indicating an evolutionary relationship between these two sequences, but rather is looking at the similarity or relatedness between their nucleic acid sequences. Many of the methods for determining

homology between two evolutionarily related molecules are routinely applied to any two or more nucleic acids or proteins for the purpose of measuring sequence similarity regardless of whether they are evolutionarily related or not.

185. In general, it is understood that one way to define any known variants and derivatives or those that might arise, of the disclosed genes and proteins herein, is through defining the variants and derivatives in terms of homology to specific known sequences. This identity of particular sequences disclosed herein is also discussed elsewhere herein. In general, variants of genes and proteins herein disclosed typically have at least, about 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99 percent homology to the stated sequence or the native sequence. Those of skill in the art readily understand how to determine the homology of two proteins or nucleic acids, such as genes. For example, the homology can be calculated after aligning the two sequences so that the homology is at its highest level.

186. Another way of calculating homology can be performed by published algorithms. Optimal alignment of sequences for comparison may be conducted by the local homology algorithm of Smith and Waterman *Adv. Appl. Math.* 2: 482 (1981), by the homology alignment algorithm of Needleman and Wunsch, *J. Mol. Biol.* 48: 443 (1970), by the search for similarity method of Pearson and Lipman, *Proc. Natl. Acad. Sci. U.S.A.* 85: 2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), or by inspection.

187. The same types of homology can be obtained for nucleic acids by for example the algorithms disclosed in Zuker, M. *Science* 244:48-52, 1989, Jaeger et al. *Proc. Natl. Acad. Sci. USA* 86:7706-7710, 1989, Jaeger et al. *Methods Enzymol.* 183:281-306, 1989 which are herein incorporated by reference for at least material related to nucleic acid alignment. It is understood that any of the methods typically can be used and that in certain instances the results of these various methods may differ, but the skilled artisan understands if identity is found with at least one of these methods, the sequences would be said to have the stated identity, and be disclosed herein.

188. For example, as used herein, a sequence recited as having a particular percent homology to another sequence refers to sequences that have the recited homology as calculated by any one or more of the calculation methods described above. For example, a first sequence has 80 percent homology, as defined herein, to a second sequence if the first sequence is

calculated to have 80 percent homology to the second sequence using the Zuker calculation method even if the first sequence does not have 80 percent homology to the second sequence as calculated by any of the other calculation methods. As another example, a first sequence has 80 percent homology, as defined herein, to a second sequence if the first sequence is calculated to have 80 percent homology to the second sequence using both the Zuker calculation method and the Pearson and Lipman calculation method even if the first sequence does not have 80 percent homology to the second sequence as calculated by the Smith and Waterman calculation method, the Needleman and Wunsch calculation method, the Jaeger calculation methods, or any of the other calculation methods. As yet another example, a first sequence has 80 percent homology, as defined herein, to a second sequence if the first sequence is calculated to have 80 percent homology to the second sequence using each of calculation methods (although, in practice, the different calculation methods will often result in different calculated homology percentages).

## 2. Hybridization/selective hybridization

189. The term hybridization typically means a sequence driven interaction between at least two nucleic acid molecules, such as a primer or a probe and a gene. Sequence driven interaction means an interaction that occurs between two nucleotides or nucleotide analogs or nucleotide derivatives in a nucleotide specific manner. For example, G interacting with C or A interacting with T are sequence driven interactions. Typically sequence driven interactions occur on the Watson-Crick face or Hoogsteen face of the nucleotide. The hybridization of two nucleic acids is affected by a number of conditions and parameters known to those of skill in the art. For example, the salt concentrations, pH, and temperature of the reaction all affect whether two nucleic acid molecules will hybridize.

190. Parameters for selective hybridization between two nucleic acid molecules are well known to those of skill in the art. For example, in some embodiments selective hybridization conditions can be defined as stringent hybridization conditions. For example, stringency of hybridization is controlled by both temperature and salt concentration of either or both of the hybridization and washing steps. For example, the conditions of hybridization to achieve selective hybridization may involve hybridization in high ionic strength solution (6X SSC or 6X SSPE) at a temperature that is about 12-25°C below the  $T_m$  (the melting temperature at which half of the molecules dissociate from their hybridization partners) followed by washing at a combination of temperature and salt concentration chosen so that the washing temperature is about 5°C to 20°C below the  $T_m$ . The temperature and salt conditions are readily determined empirically in preliminary experiments in which samples of reference DNA immobilized on

filters are hybridized to a labeled nucleic acid of interest and then washed under conditions of different stringencies. Hybridization temperatures are typically higher for DNA-RNA and RNA-RNA hybridizations. The conditions can be used as described above to achieve stringency, or as is known in the art. (Sambrook et al., *Molecular Cloning: A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1989; Kunkel et al. *Methods Enzymol.* 1987:154:367, 1987 which is herein incorporated by reference for material at least related to hybridization of nucleic acids). A preferable stringent hybridization condition for a DNA:DNA hybridization can be at about 68°C (in aqueous solution) in 6X SSC or 6X SSPE followed by washing at 68°C. Stringency of hybridization and washing, if desired, can be reduced accordingly as the degree of complementarity desired is decreased, and further, depending upon the G-C or A-T richness of any area wherein variability is searched for. Likewise, stringency of hybridization and washing, if desired, can be increased accordingly as homology desired is increased, and further, depending upon the G-C or A-T richness of any area wherein high homology is desired, all as known in the art.

191. Another way to define selective hybridization is by looking at the amount (percentage) of one of the nucleic acids bound to the other nucleic acid. For example, in some embodiments selective hybridization conditions would be when at least about, 60, 65, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100 percent of the limiting nucleic acid is bound to the non-limiting nucleic acid. Typically, the non-limiting primer is in for example, 10 or 100 or 1000 fold excess. This type of assay can be performed at under conditions where both the limiting and non-limiting primer are for example, 10 fold or 100 fold or 1000 fold below their  $k_d$ , or where only one of the nucleic acid molecules is 10 fold or 100 fold or 1000 fold or where one or both nucleic acid molecules are above their  $k_d$ .

192. Another way to define selective hybridization is by looking at the percentage of primer that gets enzymatically manipulated under conditions where hybridization is required to promote the desired enzymatic manipulation. For example, in some embodiments selective hybridization conditions would be when at least about, 60, 65, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100 percent of the primer is enzymatically manipulated under conditions which promote the enzymatic manipulation, for example if the enzymatic manipulation is DNA extension, then selective hybridization conditions would be when at least about 60, 65, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100 percent of the

primer molecules are extended. Preferred conditions also include those suggested by the manufacturer or indicated in the art as being appropriate for the enzyme performing the manipulation.

193. Just as with homology, it is understood that there are a variety of methods herein disclosed for determining the level of hybridization between two nucleic acid molecules. It is understood that these methods and conditions may provide different percentages of hybridization between two nucleic acid molecules, but unless otherwise indicated meeting the parameters of any of the methods would be sufficient. For example if 80% hybridization was required and as long as hybridization occurs within the required parameters in any one of these methods it is considered disclosed herein.

194. It is understood that those of skill in the art understand that if a composition or method meets any one of these criteria for determining hybridization either collectively or singly it is a composition or method that is disclosed herein.

### 3. Nucleic acids

195. There are a variety of molecules disclosed herein that are nucleic acid based, including for example the nucleic acids that encode, for example DHR96 or variants or fragments thereof, as well as various functional nucleic acids. The disclosed nucleic acids are made up of for example, nucleotides, nucleotide analogs, or nucleotide substitutes. Non-limiting examples of these and other molecules are discussed herein. It is understood that for example, when a vector is expressed in a cell, that the expressed mRNA will typically be made up of A, C, G, and U. Likewise, it is understood that if, for example, an antisense molecule is introduced into a cell or cell environment through for example exogenous delivery, it is advantageous that the antisense molecule be made up of nucleotide analogs that reduce the degradation of the antisense molecule in the cellular environment.

#### a) Nucleotides and related molecules

196. A nucleotide is a molecule that contains a base moiety, a sugar moiety and a phosphate moiety. Nucleotides can be linked together through their phosphate moieties and sugar moieties creating an internucleoside linkage. The base moiety of a nucleotide can be adenin-9-yl (A), cytosin-1-yl (C), guanin-9-yl (G), uracil-1-yl (U), and thymine-1-yl (T). The sugar moiety of a nucleotide is a ribose or a deoxyribose. The phosphate moiety of a nucleotide is pentavalent phosphate. A non-limiting example of a nucleotide would be 3'-AMP (3'-adenosine monophosphate) or 5'-GMP (5'-guanosine monophosphate).

197. A nucleotide analog is a nucleotide which contains some type of modification to either the base, sugar, or phosphate moieties. Modifications to the base moiety would include natural and synthetic modifications of A, C, G, and T/U as well as different purine or pyrimidine bases, such as uracil-5-yl (.psi.), hypoxanthin-9-yl (I), and 2-aminoadenin-9-yl. A modified base  
5 includes but is not limited to 5-methylcytosine (5-me-C), 5-hydroxymethyl cytosine, xanthine, hypoxanthine, 2-aminoadenine, 6-methyl and other alkyl derivatives of adenine and guanine, 2-propyl and other alkyl derivatives of adenine and guanine, 2-thiouracil, 2-thiothymine and

198. 2-thiocytosine, 5-halouracil and cytosine, 5-propynyl uracil and cytosine, 6-azo uracil, cytosine and thymine, 5-uracil (pseudouracil), 4-thiouracil, 8-halo, 8-amino, 8-thiol,  
10 8-thioalkyl, 8-hydroxyl and other 8-substituted adenines and guanines, 5-halo particularly 5-bromo, 5-trifluoromethyl and other 5-substituted uracils and cytosines, 7-methylguanine and 7-methyladenine, 8-azaguanine and 8-azaadenine, 7-deazaguanine and 7-deazaadenine and 3-deazaguanine and 3-deazaadenine. Additional base modifications can be found for example in U.S. Pat. No. 3,687,808, Englisch et al., Angewandte Chemie, International Edition, 1991, 30,  
15 613, and Sanghvi, Y. S., Chapter 15, Antisense Research and Applications, pages 289-302, Crooke, S. T. and Lebleu, B. ed., CRC Press, 1993. Certain nucleotide analogs, such as 5-substituted pyrimidines, 6-azapyrimidines and N-2, N-6 and O-6 substituted purines, including 2-aminopropyladenine, 5-propynyluracil and 5-propynylcytosine. 5-methylcytosine can increase the stability of duplex formation. Often time base modifications can be combined with for  
20 example a sugar modification, such as 2'-O-methoxyethyl, to achieve unique properties such as increased duplex stability. There are numerous United States patents such as 4,845,205; 5,130,302; 5,134,066; 5,175,273; 5,367,066; 5,432,272; 5,457,187; 5,459,255; 5,484,908; 5,502,177; 5,525,711; 5,552,540; 5,587,469; 5,594,121, 5,596,091; 5,614,617; and 5,681,941, which detail and describe a range of base modifications. Each of these patents is herein  
25 incorporated by reference.

199. Nucleotide analogs can also include modifications of the sugar moiety. Modifications to the sugar moiety would include natural modifications of the ribose and deoxy ribose as well as synthetic modifications. Sugar modifications include but are not limited to the following modifications at the 2' position: OH; F; O-, S-, or N-alkyl; O-, S-, or N-alkenyl; O-, S- or N-alkynyl; or O-alkyl-O-alkyl, wherein the alkyl, alkenyl and alkynyl may be substituted or  
30 unsubstituted C<sub>1</sub> to C<sub>10</sub>, alkyl or C<sub>2</sub> to C<sub>10</sub> alkenyl and alkynyl. 2' sugar modifications also include but are not limited to -O[(CH<sub>2</sub>)<sub>n</sub> O]<sub>m</sub> CH<sub>3</sub>, -O(CH<sub>2</sub>)<sub>n</sub> OCH<sub>3</sub>, -O(CH<sub>2</sub>)<sub>n</sub> NH<sub>2</sub>, -O(CH<sub>2</sub>)<sub>n</sub> CH<sub>3</sub>, -O(CH<sub>2</sub>)<sub>n</sub> -ONH<sub>2</sub>, and -O(CH<sub>2</sub>)<sub>n</sub> ON[(CH<sub>2</sub>)<sub>n</sub> CH<sub>3</sub>]<sub>2</sub>, where n and m are from 1 to about 10.

200. Other modifications at the 2' position include but are not limited to: C<sub>1</sub> to C<sub>10</sub> lower alkyl, substituted lower alkyl, alkaryl, aralkyl, O-alkaryl or O-aralkyl, SH, SCH<sub>3</sub>, OCN, Cl, Br, CN, CF<sub>3</sub>, OCF<sub>3</sub>, SOCH<sub>3</sub>, SO<sub>2</sub> CH<sub>3</sub>, ONO<sub>2</sub>, NO<sub>2</sub>, N<sub>3</sub>, NH<sub>2</sub>, heterocycloalkyl, heterocycloalkaryl, aminoalkylamino, polyalkylamino, substituted silyl, an RNA cleaving group, a reporter group, an intercalator, a group for improving the pharmacokinetic properties of an oligonucleotide, or a group for improving the pharmacodynamic properties of an oligonucleotide, and other substituents having similar properties. Similar modifications may also be made at other positions on the sugar, particularly the 3' position of the sugar on the 3' terminal nucleotide or in 2'-5' linked oligonucleotides and the 5' position of 5' terminal nucleotide. Modified sugars would also include those that contain modifications at the bridging ring oxygen, such as CH<sub>2</sub> and S. Nucleotide sugar analogs may also have sugar mimetics such as cyclobutyl moieties in place of the pentofuranosyl sugar. There are numerous United States patents that teach the preparation of such modified sugar structures such as 4,981,957; 5,118,800; 5,319,080; 5,359,044; 5,393,878; 5,446,137; 5,466,786; 5,514,785; 5,519,134; 5,567,811; 5,576,427; 5,591,722; 5,597,909; 5,610,300; 5,627,053; 5,639,873; 5,646,265; 5,658,873; 5,670,633; and 5,700,920, each of which is herein incorporated by reference in its entirety.

201. Nucleotide analogs can also be modified at the phosphate moiety. Modified phosphate moieties include but are not limited to those that can be modified so that the linkage between two nucleotides contains a phosphorothioate, chiral phosphorothioate, phosphorodithioate, phosphotriester, aminoalkylphosphotriester, methyl and other alkyl phosphonates including 3'-alkylene phosphonate and chiral phosphonates, phosphinates, phosphoramidates including 3'-amino phosphoramidate and aminoalkylphosphoramidates, thionophosphoramidates, thionoalkylphosphonates, thionoalkylphosphotriesters, and boranophosphates. It is understood that these phosphate or modified phosphate linkage between two nucleotides can be through a 3'-5' linkage or a 2'-5' linkage, and the linkage can contain inverted polarity such as 3'-5' to 5'-3' or 2'-5' to 5'-2'. Various salts, mixed salts and free acid forms are also included. Numerous United States patents teach how to make and use nucleotides containing modified phosphates and include but are not limited to, 3,687,808; 4,469,863; 4,476,301; 5,023,243; 5,177,196; 5,188,897; 5,264,423; 5,276,019; 5,278,302; 5,286,717; 5,321,131; 5,399,676; 5,405,939; 5,453,496; 5,455,233; 5,466,677; 5,476,925; 5,519,126; 5,536,821; 5,541,306; 5,550,111; 5,563,253; 5,571,799; 5,587,361; and 5,625,050, each of which is herein incorporated by reference.

202. It is understood that nucleotide analogs need only contain a single modification, but may also contain multiple modifications within one of the moieties or between different moieties.

203. Nucleotide substitutes are molecules having similar functional properties to nucleotides, but which do not contain a phosphate moiety, such as peptide nucleic acid (PNA). Nucleotide substitutes are molecules that will recognize nucleic acids in a Watson-Crick or Hoogsteen manner, but which are linked together through a moiety other than a phosphate moiety. Nucleotide substitutes are able to conform to a double helix type structure when interacting with the appropriate target nucleic acid.

204. Nucleotide substitutes are nucleotides or nucleotide analogs that have had the phosphate moiety and/or sugar moieties replaced. Nucleotide substitutes do not contain a standard phosphorus atom. Substitutes for the phosphate can be for example, short chain alkyl or cycloalkyl internucleoside linkages, mixed heteroatom and alkyl or cycloalkyl internucleoside linkages, or one or more short chain heteroatomic or heterocyclic internucleoside linkages. These include those having morpholino linkages (formed in part from the sugar portion of a nucleoside); siloxane backbones; sulfide, sulfoxide and sulfone backbones; formacetyl and thioformacetyl backbones; methylene formacetyl and thioformacetyl backbones; alkene containing backbones; sulfamate backbones; methyleneimino and methylenehydrazino backbones; sulfonate and sulfonamide backbones; amide backbones; and others having mixed N, O, S and CH<sub>2</sub> component parts. Numerous United States patents disclose how to make and use these types of phosphate replacements and include but are not limited to 5,034,506; 5,166,315; 5,185,444; 5,214,134; 5,216,141; 5,235,033; 5,264,562; 5,264,564; 5,405,938; 5,434,257; 5,466,677; 5,470,967; 5,489,677; 5,541,307; 5,561,225; 5,596,086; 5,602,240; 5,610,289; 5,602,240; 5,608,046; 5,610,289; 5,618,704; 5,623,070; 5,663,312; 5,633,360; 5,677,437; and 5,677,439, each of which is herein incorporated by reference.

205. It is also understood in a nucleotide substitute that both the sugar and the phosphate moieties of the nucleotide can be replaced, by for example an amide type linkage (aminoethylglycine) (PNA). United States patents 5,539,082; 5,714,331; and 5,719,262 teach how to make and use PNA molecules, each of which is herein incorporated by reference. (See also Nielsen et al., Science, 1991, 254, 1497-1500).

206. It is also possible to link other types of molecules (conjugates) to nucleotides or nucleotide analogs to enhance for example, cellular uptake. Conjugates can be chemically linked

to the nucleotide or nucleotide analogs. Such conjugates include but are not limited to lipid moieties such as a cholesterol moiety (Letsinger et al., Proc. Natl. Acad. Sci. USA, 1989,

207. 86, 6553-6556), cholic acid (Manoharan et al., Bioorg. Med. Chem. Lett., 1994, 4, 1053-1060), a thioether, e.g., hexyl-S-tritylthiol (Manoharan et al., Ann. N.Y. Acad. Sci., 1992, 660, 306-309; Manoharan et al., Bioorg. Med. Chem. Lett., 1993, 3, 2765-2770), a thiocholesterol (Oberhauser et al., Nucl. Acids Res., 1992, 20, 533-538), an aliphatic chain, e.g., dodecandiol or undecyl residues (Saison-Behmoaras et al., EMBO J., 1991, 10, 1111-1118; Kabanov et al., FEBS Lett., 1990, 259, 327-330; Svinarchuk et al., Biochimie, 1993, 75, 49-54), a phospholipid, e.g., di-hexadecyl-rac-glycerol or triethylammonium

1,2-di-O-hexadecyl-rac-glycero-3-H-phosphonate (Manoharan et al., Tetrahedron Lett., 1995, 36, 3651-3654; Shea et al., Nucl. Acids Res., 1990, 18, 3777-3783), a polyamine or a polyethylene glycol chain (Manoharan et al., Nucleosides & Nucleotides, 1995, 14, 969-973), or adamantane acetic acid (Manoharan et al., Tetrahedron Lett., 1995, 36, 3651-3654), a palmityl moiety (Mishra et al., Biochim. Biophys. Acta, 1995, 1264, 229-237), or an octadecylamine or hexylamino-carbonyl-oxycholesterol moiety (Crooke et al., J. Pharmacol. Exp. Ther., 1996, 277, 923-937. Numerous United States patents teach the preparation of such conjugates and include, but are not limited to U.S. Pat. Nos. 4,828,979; 4,948,882; 5,218,105; 5,525,465; 5,541,313; 5,545,730; 5,552,538; 5,578,717; 5,580,731; 5,580,731; 5,591,584; 5,109,124; 5,118,802; 5,138,045; 5,414,077; 5,486,603; 5,512,439; 5,578,718; 5,608,046; 4,587,044; 4,605,735; 4,667,025; 4,762,779; 4,789,737; 4,824,941; 4,835,263; 4,876,335; 4,904,582; 4,958,013; 5,082,830; 5,112,963; 5,214,136; 5,082,830; 5,112,963; 5,214,136; 5,245,022; 5,254,469; 5,258,506; 5,262,536; 5,272,250; 5,292,873; 5,317,098; 5,371,241; 5,391,723; 5,416,203; 5,451,463; 5,510,475; 5,512,667; 5,514,785; 5,565,552; 5,567,810; 5,574,142; 5,585,481; 5,587,371; 5,595,726; 5,597,696; 5,599,923; 5,599,928 and 5,688,941, each of which is herein incorporated by reference.

208. A Watson-Crick interaction is at least one interaction with the Watson-Crick face of a nucleotide, nucleotide analog, or nucleotide substitute. The Watson-Crick face of a nucleotide, nucleotide analog, or nucleotide substitute includes the C2, N1, and C6 positions of a purine based nucleotide, nucleotide analog, or nucleotide substitute and the C2, N3, C4 positions of a pyrimidine based nucleotide, nucleotide analog, or nucleotide substitute.

209. A Hoogsteen interaction is the interaction that takes place on the Hoogsteen face of a nucleotide or nucleotide analog, which is exposed in the major groove of duplex DNA. The

Hoogsteen face includes the N7 position and reactive groups (NH<sub>2</sub> or O) at the C6 position of purine nucleotides.

### **b) Sequences**

210. There are a variety of sequences related to the DHR96 gene, and these sequences  
5 and others are herein incorporated by reference in their entireties as well as for individual subsequences contained therein.

211. One particular sequence set forth in SEQ ID NO:7 and having Genbank accession  
number NM\_079769 is used herein, as an example, to exemplify the disclosed compositions and  
methods. It is understood that the description related to this sequence is applicable to any  
10 sequence related to DHR96 or any other sequences disclosed herein, unless specifically indicated  
otherwise. Those of skill in the art understand how to resolve sequence discrepancies and  
differences and to adjust the compositions and methods relating to a particular sequence to other  
related sequences (i.e. sequences of DHR96 or variants or fragments thereof). Primers and/or  
probes can be designed for any DHR96 sequence given the information disclosed herein and  
15 known in the art.

### **c) Primers and probes**

212. Disclosed are compositions including primers and probes, which are capable of  
interacting with the genes disclosed herein. In certain embodiments the primers are used to  
support DNA amplification reactions. Typically the primers will be capable of being extended in  
20 a sequence specific manner. Extension of a primer in a sequence specific manner includes any  
methods wherein the sequence and/or composition of the nucleic acid molecule to which the  
primer is hybridized or otherwise associated directs or influences the composition or sequence of  
the product produced by the extension of the primer. Extension of the primer in a sequence  
specific manner therefore includes, but is not limited to, PCR, DNA sequencing, DNA  
25 extension, DNA polymerization, RNA transcription, or reverse transcription. Techniques and  
conditions that amplify the primer in a sequence specific manner are preferred. In certain  
embodiments the primers are used for the DNA amplification reactions, such as PCR or direct  
sequencing. It is understood that in certain embodiments the primers can also be extended using  
non-enzymatic techniques, where for example, the nucleotides or oligonucleotides used to  
30 extend the primer are modified such that they will chemically react to extend the primer in a  
sequence specific manner. Typically the disclosed primers hybridize with the nucleic acid or  
region of the nucleic acid or they hybridize with the complement of the nucleic acid or  
complement of a region of the nucleic acid.

#### 4. Delivery of the compositions to cells

213. There are a number of compositions and methods which can be used to deliver nucleic acids to cells, either in vitro or in vivo. These methods and compositions can largely be broken down into two classes: viral based delivery systems and non-viral based delivery systems.

5 For example, the nucleic acids can be delivered through a number of direct delivery systems such as, electroporation, lipofection, calcium phosphate precipitation, plasmids, viral vectors, viral nucleic acids, phage nucleic acids, phages, cosmids, or via transfer of genetic material in cells or carriers such as cationic liposomes. Appropriate means for transfection, including viral vectors, chemical transfectants, or physico-mechanical methods such as electroporation and  
10 direct diffusion of DNA, are described by, for example, Wolff, J. A., et al., Science, 247, 1465-1468, (1990); and Wolff, J. A. Nature, 352, 815-818, (1991) Such methods are well known in the art and readily adaptable for use with the compositions and methods described herein. In certain cases, the methods will be modified to specifically function with large DNA molecules. Further, these methods can be used to target certain diseases and cell populations by using the  
15 targeting characteristics of the carrier.

##### a) Nucleic acid based delivery systems

214. The term "transgene" is used herein to describe genetic material which is artificially inserted into the genome of an invertebrate cell. The transgene encodes a product that, when expressed in embryos, gives rise to a specific phenotype. A transgene can encode a  
20 transcription factor or mimetic thereof having the desired result. A recombinant DNA molecule or vector containing a heterologous protein gene expression unit can be used to transfect invertebrate cells (United States Patents 4,670,388 and 5,550,043, herein incorporated by reference in their entirety.) A gene expression unit can contain a DNA coding sequence for a selected protein or for a derivative thereof. Such derivatives can be obtained by manipulation of  
25 the gene sequence using traditional genetic engineering techniques, e.g., mutagenesis, restriction endonuclease treatment, ligation of other gene sequences including synthetic sequences and the like (T. Maniatis et al, Molecular Cloning, A Laboratory Manual., Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. (1982).

215. Expression of the transgene can be targeted to occur in a non-adult stage of the  
30 animal, the transgene can be stably integrated into the genome of the animal in a manner such that its expression is controlled both spatially and temporally to the desired cell type and the correct developmental stage, i.e. to expression in embryonic neuroblasts. Specifically, the subject transgene can stably integrated into the genome of the animal under the control of a promoter

that provides for expression. The transgene may be under the control of any convenient promoter that provides for this requisite spatial and temporal expression pattern, where the promoter can be endogenous or exogenous. A suitable promoter is the promoter located in the *Drosophila melanogaster* genome at position 86E1-3.

216. Another suitable promoter of the *Drosophila* origin includes the *Drosophila* metallothionein promoter (Lastowski-Perry et al, J. Biol. Chem., 260:1527, 1985). This inducible promoter directs high-level transcription of the gene in the presence of metals, e.g., CuSO<sub>4</sub>. Use of the *Drosophila* metallothionein promoter results in the expression system of the invention retaining full regulation even at very high copy number. This is in direct contrast to the use of the mammalian metallothionein promoter in mammalian cells in which the regulatory effect of the metal is diminished as copy number increases. In the *Drosophila* expression system, this retained inducibility effect increases expression of the gene product in the *Drosophila* cell at high copy number.

217. The *Drosophila* actin 5C gene promoter (B. J. Bond et al, Mol. Cell. Biol., 6: 2080, 1986) is also a desirable promoter sequence. The actin 5C promoter is a constitutive promoter and does not require addition of metal. Therefore, it is better-suited for use in a large scale production system, like a perfusion system, than is the *Drosophila* metallothionein promoter. An additional advantage is that the absence of a high concentration of copper in the media maintains the cells in a healthier state for longer periods of time.

218. Examples of other known *Drosophila* promoters include, e.g., the inducible heatshock (Hsp70) and COPIA LTR promoters. The SV40 early promoter gives lower levels of expression than the *Drosophila* metallothionein promoter.

219. The transgene may be integrated into the fly genome in a manner that provides for direct or indirect expression activation by the promoter, i.e. in a manner that provides for either cis or trans activation of gene expression by the promoter. In other words, expression of the transgene may be mediated directly by the promoter, or through one or more transactivating agents. Where the transgene is under direct control of the promoter, i.e. the promoter regulates expression of the transgene in a cis fashion, the transgene is stably integrated into the genome of the fly at a site sufficiently proximal to the promoter and in frame with the promoter such that cis regulation by the promoter occurs.

220. In other embodiments where expression of the transgene is indirectly mediated by the endogenous promoter, the promoter controls expression of the transgene through one or more transactivating agents, usually one transactivating agent, i.e. an agent whose expression is

directly controlled by the promoter and which binds to the region of the transgene in a manner sufficient to turn on expression of the transgene. Any convenient transactivator may be employed. The GAL4 transactivator system an example of such a system.

221. The GAL4 encoding sequence can be stably integrated into the genome of the animal in a manner such that it is operatively linked to the endogenous promoter that provides expression in the appropriate location. The GAL4 system consists of the yeast transcriptional activator GAL4 and its target the upstream activating sequence (UAS) located within the P-element. Initially, GAL4 and UAS are in separate lines. The UAS is mobilized to generate new UAS insertion lines which remain silent until a source of GAL4 is made available. Under the control of a promoter, the expression of GAL4 is directed in a particular pattern. Specialized promoters can be used to drive expression of GAL4 in tissue and cell specific manners. The GAL4 containing line is then crossed to the UAS containing line. The UAS in the presence of GAL4 directs the expression of any genes adjacent to its insertion site. When the insertion site is located upstream from the coding region over-or ectopic expression occurs.

222. Flies of line 31-1 (also referred to as 1822), as disclosed in Brand & Perrimon, Development (1993) 118: 401-415 express GAL4 in this manner, and are known to those of skill in the art. The transgene is stably integrated into a different location of the genome, generally a random location in the genome, where the transgene is operatively linked to an upstream activator sequence, i.e. UAS sequence, to which GAL4 binds and turns on expression of the transgene. Transgenic flies having a UAS: GAL4 transactivation system are known to those of skill in the art and are described in Brand & Perrimon, Development (1993) 118: 401-415; and Phelps & Brand, Methods (April 1998) 14:367-379.

223. A desirable gene expression unit or expression vector for the protein of interest can also be constructed by fusing the protein coding sequence to a desirable signal sequence. The signal sequence functions to direct secretion of the protein from the host cell. Such a signal sequence may be derived from the sequence of tissue plasminogen activator (tPA). Other available signal sequences include, e.g., those derived from Herpes Simplex virus gene HSV-I gD (Lasky et al, Science, 233:209-212 1986).

224. The DNA coding sequence can also be followed by a polyadenylation (poly A) region, such as an SV40 early poly A region. The poly A region which functions in the polyadenylation of RNA transcripts appears to play a role in stabilizing transcription. A similar poly A region can be derived from a variety of genes in which it is naturally present. This region

can also be modified to alter its sequence provided that polyadenylation and transcript stabilization functions are not significantly adversely affected.

225. The recombinant DNA molecule may also carry a genetic selection marker, as well as the protein gene functions. The selection marker can be any gene or genes which cause a readily detectable phenotypic change in a transfected host cell. Such phenotypic change can be, for example, drug resistance, such as the gene for hygromycin B resistance (i.e., hygromycin B phosphotransferase).

226. Alternatively, a selection system using the drug methotrexate, and prokaryotic dihydrofolate reductase (DHFR) gene, can be used with Invertebrate cells. The endogenous eukaryotic DHFR of the cells is inhibited by methotrexate. Therefore, by transfecting the cells with a plasmid containing the prokaryotic DHFR which is insensitive to methotrexate and selecting with methotrexate, only cells transfected with and expressing the prokaryotic DHFR will survive. Unlike methotrexate, selection of transformed mammalian and bacterial cells, in the *Drosophila* system, methotrexate can be used to initially high-copy number transfectants. Only cells which have incorporated the protective prokaryotic DHFR gene will survive. Concomitantly, these cells have the gene expression unit of interest.

227. The subject transgenic flies can be prepared using any convenient protocol that provides for stable integration of the transgene into the fly genome in a manner sufficient to provide for the requisite spatial and temporal expression of the transgene, i.e. in embryonic neuroblasts. A number of different strategies can be employed to obtain the integration of the transgene with the requisite expression pattern. Generally, methods of producing the subject transgenic flies involve stable integration of the transgene into the fly genome. Stable integration is achieved by first introducing the transgene into a cell or cells of the fly, e.g. a fly embryo. The transgene is generally present on a suitable vector, such as a plasmid. Transgene introduction may be accomplished using any convenient protocol, where suitable protocols include: electroporation, microinjection, vesicle delivery, e.g. liposome delivery vehicles, and the like. Following introduction of the transgene into the cell(s), the transgene is stably integrated into the genome of the cell. Stable integration may be either site specific or random, but is generally random.

228. Where integration is random, the transgene is typically integrated with the use of transposase. In such embodiments, the transgene can be introduced into the cell(s) within a vector that includes the requisite P element, terminal 31 base pair inverted repeats. Where the cell into which the transgene is to be integrated does not comprise an endogenous transposase, a

vector encoding a transposase can also be introduced into the cell, e.g. a helper plasmid comprising a transposase gene, such as pTURBO (Steller & Pirrotta, Mol. Cell. Biol. 6:1640-1649, 1986). Methods of random integration of transgenes into the genome of a target *Drosophila melanogaster* cell(s) are disclosed in U.S. Pat. No. 4,670,388, the disclosure of which is herein incorporated by reference.

229. Transcription and expression of the heterologous protein coding sequences can be monitored. For example, Southern blot analysis can be used to determine copy number of the gp120 gene. Northern blot analysis provides information regarding the size of the transcribed gene sequence. The level of transcription can also be quantitated. Expression of the selected protein in the recombinant cells can be further verified through Western blot analysis, for example.

230. In those embodiments in which the transgene is stably integrated in a random fashion into the fly genome, means are also provided for selectively expressing the transgene at the appropriate time during development of the fly. In other words, means are provided for obtaining targeted expression of the transgene. To obtain the desired targeted expression of the randomly integrated transgene, integration of particular promoter upstream of the transgene, as a single unit in the P element vector may be employed. Alternatively, a transactivator that mediates expression of the transgene may be employed. Of particular interest is the GAL4 system described in Brand & Perrimon, Development (1993) 118: 401-415; and Phelps & Brand, Methods (April 1998) 14:367-379.

231. In one embodiment, the subject transgenic flies are produced by: (1) generating two separate lines of transgenic flies: (a) a first line that expresses GAL4; and (b) a second line in which the transgene is stably integrated into the cell genome and is fused to a UAS domain; (2) crossing the two lines; and (3) screening the progeny for the desired phenotype, i.e. adult onset neurodegeneration. Each of the above steps are well known to those of skill in the art (Brand & Perrimon, Development 118: 401-415, 1993; and Phelps & Brand, Methods 14:367-379, April 1998.)

#### **b) Non-nucleic acid based systems**

232. The disclosed compositions can be delivered to the target cells in a variety of ways. For example, the compositions can be delivered through electroporation, or through lipofection, or through calcium phosphate precipitation. The delivery mechanism chosen will depend in part on the type of cell targeted and whether the delivery is occurring for example in vivo or in vitro.

233. Thus, the compositions can comprise, in addition to the disclosed compositions or vectors for example, lipids such as liposomes, such as cationic liposomes (e.g., DOTMA, DOPE, DC-cholesterol) or anionic liposomes. Liposomes can further comprise proteins to facilitate targeting a particular cell, if desired. Administration of a composition comprising a compound and a cationic liposome can be administered to the blood afferent to a target organ or inhaled into the respiratory tract to target cells of the respiratory tract. Regarding liposomes, see, e.g., Brigham et al. *Am. J. Resp. Cell. Mol. Biol.* 1:95-100 (1989); Felgner et al. *Proc. Natl. Acad. Sci USA* 84:7413-7417 (1987); U.S. Pat. No.4,897,355. Furthermore, the compound can be administered as a component of a microcapsule that can be targeted to specific cell types, such as macrophages, or where the diffusion of the compound or delivery of the compound from the microcapsule is designed for a specific rate or dosage.

234. In the methods described above which include the administration and uptake of exogenous DNA into the cells of a subject (i.e., gene transduction or transfection), delivery of the compositions to cells can be via a variety of mechanisms. As one example, delivery can be via a liposome, using commercially available liposome preparations such as LIPOFECTIN, LIPOFECTAMINE (GIBCO-BRL, Inc., Gaithersburg, MD), SUPERFECT (Qiagen, Inc. Hilden, Germany) and TRANSFECTAM (Promega Biotec, Inc., Madison, WI), as well as other liposomes developed according to procedures standard in the art. In addition, the disclosed nucleic acid or vector can be delivered *in vivo* by electroporation, the technology for which is available from Genetronics, Inc. (San Diego, CA) as well as by means of a SONOPORATION machine (ImaRx Pharmaceutical Corp., Tucson, AZ).

235. The materials may be in solution, suspension (for example, incorporated into microparticles, liposomes, or cells). These may be targeted to a particular cell type via antibodies, receptors, or receptor ligands. The following references are examples of the use of this technology to target specific proteins to tumor tissue (Senter, et al., Bioconjugate Chem., 2:447-451, (1991); Bagshawe, K.D., Br. J. Cancer, 60:275-281, (1989); Bagshawe, et al., Br. J. Cancer, 58:700-703, (1988); Senter, et al., Bioconjugate Chem., 4:3-9, (1993); Battelli, et al., Cancer Immunol. Immunother., 35:421-425, (1992); Pietersz and McKenzie, Immunolog. Reviews, 129:57-80, (1992); and Roffler, et al., Biochem. Pharmacol., 42:2062-2065, (1991)). These techniques can be used for a variety of other specific cell types. Vehicles such as "stealth" and other antibody conjugated liposomes (including lipid mediated drug targeting to colonic carcinoma), receptor mediated targeting of DNA through cell specific ligands, lymphocyte directed tumor targeting, and highly specific therapeutic retroviral targeting of murine glioma

cells *in vivo*. The following references are examples of the use of this technology to target specific proteins to tumor tissue (Hughes et al., Cancer Research, 49:6214-6220, (1989); and Litzinger and Huang, Biochimica et Biophysica Acta, 1104:179-187, (1992)). In general, receptors are involved in pathways of endocytosis, either constitutive or ligand induced. These receptors cluster in clathrin-coated pits, enter the cell via clathrin-coated vesicles, pass through an acidified endosome in which the receptors are sorted, and then either recycle to the cell surface, become stored intracellularly, or are degraded in lysosomes. The internalization pathways serve a variety of functions, such as nutrient uptake, removal of activated proteins, clearance of macromolecules, opportunistic entry of viruses and toxins, dissociation and degradation of ligand, and receptor-level regulation. Many receptors follow more than one intracellular pathway, depending on the cell type, receptor concentration, type of ligand, ligand valency, and ligand concentration. Molecular and cellular mechanisms of receptor-mediated endocytosis has been reviewed (Brown and Greene, DNA and Cell Biology 10:6, 399-409 (1991)).

236. Nucleic acids that are delivered to cells which are to be integrated into the host cell genome, typically contain integration sequences. These sequences are often viral related sequences, particularly when viral based systems are used. These viral intergration systems can also be incorporated into nucleic acids which are to be delivered using a non-nucleic acid based system of deliver, such as a liposome, so that the nucleic acid contained in the delivery system can be come integrated into the host genome.

237. Other general techniques for integration into the host genome include, for example, systems designed to promote homologous recombination with the host genome. These systems typically rely on sequence flanking the nucleic acid to be expressed that has enough homology with a target sequence within the host cell genome that recombination between the vector nucleic acid and the target nucleic acid takes place, causing the delivered nucleic acid to be integrated into the host genome. These systems and the methods necessary to promote homologous recombination are known to those of skill in the art.

#### c) *In vivo/ex vivo*

238. As described above, the compositions can be administered in a pharmaceutically acceptable carrier and can be delivered to the subject=s cells *in vivo* and/or *ex vivo* by a variety of mechanisms well known in the art (e.g., uptake of naked DNA, liposome fusion, intramuscular injection of DNA via a gene gun, endocytosis and the like).

239. If *ex vivo* methods are employed, cells or tissues can be removed and maintained outside the body according to standard protocols well known in the art. The compositions can be introduced into the cells via any gene transfer mechanism, such as, for example, calcium phosphate mediated gene delivery, electroporation, microinjection or proteoliposomes. The transduced cells can then be infused (e.g., in a pharmaceutically acceptable carrier) or homotopically transplanted back into the subject per standard methods for the cell or tissue type. Standard methods are known for transplantation or infusion of various cells into a subject.

## 5. Peptides

### a) Protein variants

240. As discussed herein there are numerous variants of the DHR96 protein that are known and herein contemplated. In addition, to the known functional DHR96 strain variants there are derivatives of the DHR96 protein which also function in the disclosed methods and compositions. Protein variants and derivatives are well understood to those of skill in the art and in can involve amino acid sequence modifications. For example, amino acid sequence modifications typically fall into one or more of three classes: substitutional, insertional or deletional variants. Insertions include amino and/or carboxyl terminal fusions as well as intrasequence insertions of single or multiple amino acid residues. Insertions ordinarily will be smaller insertions than those of amino or carboxyl terminal fusions, for example, on the order of one to four residues. Immunogenic fusion protein derivatives, such as those described in the examples, are made by fusing a polypeptide sufficiently large to confer immunogenicity to the target sequence by cross-linking in vitro or by recombinant cell culture transformed with DNA encoding the fusion. Deletions are characterized by the removal of one or more amino acid residues from the protein sequence. Typically, no more than about from 2 to 6 residues are deleted at any one site within the protein molecule. These variants ordinarily are prepared by site specific mutagenesis of nucleotides in the DNA encoding the protein, thereby producing DNA encoding the variant, and thereafter expressing the DNA in recombinant cell culture. Techniques for making substitution mutations at predetermined sites in DNA having a known sequence are well known, for example M13 primer mutagenesis and PCR mutagenesis. Amino acid substitutions are typically of single residues, but can occur at a number of different locations at once; insertions usually will be on the order of about from 1 to 10 amino acid residues; and deletions will range about from 1 to 30 residues. Deletions or insertions preferably are made in adjacent pairs, i.e. a deletion of 2 residues or insertion of 2 residues. Substitutions, deletions, insertions or any combination thereof may be combined to arrive at a final construct. The

mutations must not place the sequence out of reading frame and preferably will not create complementary regions that could produce secondary mRNA structure. Substitutional variants are those in which at least one residue has been removed and a different residue inserted in its place. Such substitutions generally are made in accordance with the following Tables 1 and 2 and are referred to as conservative substitutions.

241. TABLE 1:Amino Acid Abbreviations

Amino Acid	Abbreviations
alanine	AlaA
allosoleucine	Alle
arginine	ArgR
asparagine	AsnN
aspartic acid	AspD
cysteine	CysC
glutamic acid	GluE
glutamine	GlnK
glycine	GlyG
histidine	HisH
isoleucine	IleI
leucine	LeuL
lysine	LysK
phenylalanine	PheF
proline	ProP
pyroglutamic acidp	Glu
serine	SerS
threonine	ThrT
tyrosine	TyrY
tryptophan	TrpW
valine	ValV

TABLE 2:Amino Acid Substitutions

Original Residue Exemplary Conservative Substitutions, others are known in the art.
Alaser
Arglys, gln
Asngln; his
Aspglu
Cysser
Glnasn, lys
Gluasp
Glypro
Hisasn;gln
Ileleu; val
Leuile; val
Lysarg; gln;
MetLeu; ile
Phemet; leu; tyr
Serthr
Thrser
Trptyr
Tyrtrp; phe
Valile; leu

242. Substantial changes in function or immunological identity are made by selecting substitutions that are less conservative than those in Table 2, i.e., selecting residues that differ more significantly in their effect on maintaining (a) the structure of the polypeptide backbone in the area of the substitution, for example as a sheet or helical conformation, (b) the charge or hydrophobicity of the molecule at the target site or (c) the bulk of the side chain. The substitutions which in general are expected to produce the greatest changes in the protein properties will be those in which (a) a hydrophilic residue, e.g. seryl or threonyl, is substituted for (or by) a hydrophobic residue, e.g. leucyl, isoleucyl, phenylalanyl, valyl or alanyl; (b) a cysteine or proline is substituted for (or by) any other residue; (c) a residue having an electropositive side chain, e.g., lysyl, arginyl, or histidyl, is substituted for (or by) an electronegative residue, e.g., glutamyl or aspartyl; or (d) a residue having a bulky side chain, e.g., phenylalanine, is substituted for (or by) one not having a side chain, e.g., glycine, in this case, (e) by increasing the number of sites for sulfation and/or glycosylation.

243. For example, the replacement of one amino acid residue with another that is biologically and/or chemically similar is known to those skilled in the art as a conservative substitution. For example, a conservative substitution would be replacing one hydrophobic residue for another, or one polar residue for another. The substitutions include combinations such as, for example, Gly, Ala; Val, Ile, Leu; Asp, Glu; Asn, Gln; Ser, Thr; Lys, Arg; and Phe, Tyr. Such conservatively substituted variations of each explicitly disclosed sequence are included within the mosaic polypeptides provided herein.

244. Substitutional or deletional mutagenesis can be employed to insert sites for N-glycosylation (Asn-X-Thr/Ser) or O-glycosylation (Ser or Thr). Deletions of cysteine or other labile residues also may be desirable. Deletions or substitutions of potential proteolysis sites, e.g. Arg, is accomplished for example by deleting one of the basic residues or substituting one by glutaminyl or histidyl residues.

245. Certain post-translational derivatizations are the result of the action of recombinant host cells on the expressed polypeptide. Glutaminyl and asparaginyl residues are frequently post-translationally deamidated to the corresponding glutamyl and asparyl residues. Alternatively, these residues are deamidated under mildly acidic conditions. Other post-translational modifications include hydroxylation of proline and lysine, phosphorylation of hydroxyl groups of seryl or threonyl residues, methylation of the o-amino groups of lysine, arginine, and histidine side chains (T.E. Creighton, Proteins: Structure and Molecular

Properties, W. H. Freeman & Co., San Francisco pp 79-86 [1983]), acetylation of the N-terminal amine and, in some instances, amidation of the C-terminal carboxyl.

246. It is understood that one way to define the variants and derivatives of the disclosed proteins herein is through defining the variants and derivatives in terms of  
5 homology/identity to specific known sequences. For example, SEQ ID NO:8 sets forth a particular sequence of DHR96 cDNA and SEQ ID NO:7 sets forth a particular sequence of a DHR96 protein. Specifically disclosed are variants of these and other proteins herein disclosed which have at least, 70% or 75% or 80% or 85% or 90% or 95% homology to the stated  
10 sequence. Those of skill in the art readily understand how to determine the homology of two proteins. For example, the homology can be calculated after aligning the two sequences so that the homology is at its highest level.

247. Another way of calculating homology can be performed by published algorithms. Optimal alignment of sequences for comparison may be conducted by the local homology  
15 algorithm of Smith and Waterman *Adv. Appl. Math.* 2: 482 (1981), by the homology alignment algorithm of Needleman and Wunsch, *J. Mol. Biol.* 48: 443 (1970), by the search for similarity method of Pearson and Lipman, *Proc. Natl. Acad. Sci. U.S.A.* 85: 2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr.,  
Madison, WI), or by inspection.

248. The same types of homology can be obtained for nucleic acids by for example the  
20 algorithms disclosed in Zuker, M. *Science* 244:48-52, 1989, Jaeger et al. *Proc. Natl. Acad. Sci. USA* 86:7706-7710, 1989, Jaeger et al. *Methods Enzymol.* 183:281-306, 1989 which are herein incorporated by reference for at least material related to nucleic acid alignment.

249. It is understood that the description of conservative mutations and homology can  
25 be combined together in any combination, such as embodiments that have at least 70% homology to a particular sequence wherein the variants are conservative mutations.

250. As this specification discusses various proteins and protein sequences it is understood that the nucleic acids that can encode those protein sequences are also disclosed. This would include all degenerate sequences related to a specific protein sequence, i.e. all  
30 nucleic acids having a sequence that encodes one particular protein sequence as well as all nucleic acids, including degenerate nucleic acids, encoding the disclosed variants and derivatives of the protein sequences. Thus, while each particular nucleic acid sequence may not be written out herein, it is understood that each and every sequence is in fact disclosed and described herein

through the disclosed protein sequence. For example, one of the many nucleic acid sequences that can encode the protein sequence set forth in SEQ ID NO:7 is set forth in SEQ ID NO:8. It is also understood that while no amino acid sequence indicates what particular DNA sequence encodes that protein within an organism, where particular variants of a disclosed protein are disclosed herein, the known nucleic acid sequence that encodes that protein in the particular organism from which that protein arises is also known and herein disclosed and described.

251. It is understood that there are numerous amino acid and peptide analogs which can be incorporated into the disclosed compositions. For example, there are numerous D amino acids or amino acids which have a different functional substituent than the amino acids shown in Table 1 and Table 2. The opposite stereo isomers of naturally occurring peptides are disclosed, as well as the stereo isomers of peptide analogs. These amino acids can readily be incorporated into polypeptide chains by charging tRNA molecules with the amino acid of choice and engineering genetic constructs that utilize, for example, amber codons, to insert the analog amino acid into a peptide chain in a site specific way (Thorson et al., *Methods in Molec. Biol.* 77:43-73 (1991), Zoller, *Current Opinion in Biotechnology*, 3:348-354 (1992); Ibba, *Biotechnology & Genetic Engineering Reviews* 13:197-216 (1995), Cahill et al., *TIBS*, 14(10):400-403 (1989); Benner, *TIB Tech*, 12:158-163 (1994); Ibba and Hennecke, *Bio/technology*, 12:678-682 (1994) all of which are herein incorporated by reference at least for material related to amino acid analogs).

252. Molecules can be produced that resemble peptides, but which are not connected via a natural peptide linkage. For example, linkages for amino acids or amino acid analogs can include  $\text{CH}_2\text{NH--}$ ,  $\text{--CH}_2\text{S--}$ ,  $\text{--CH}_2\text{--CH}_2\text{--}$ ,  $\text{--CH=CH--}$  (cis and trans),  $\text{--COCH}_2\text{--}$ ,  $\text{--CH(OH)CH}_2\text{--}$ , and  $\text{--CHH}_2\text{SO--}$  (These and others can be found in Spatola, A. F. in *Chemistry and Biochemistry of Amino Acids, Peptides, and Proteins*, B. Weinstein, eds., Marcel Dekker, New York, p. 267 (1983); Spatola, A. F., *Vega Data* (March 1983), Vol. 1, Issue 3, *Peptide Backbone Modifications* (general review); Morley, *Trends Pharm Sci* (1980) pp. 463-468; Hudson, D. et al., *Int J Pept Prot Res* 14:177-185 (1979) ( $\text{--CH}_2\text{NH--}$ ,  $\text{CH}_2\text{CH}_2\text{--}$ ); Spatola et al. *Life Sci* 38:1243-1249 (1986) ( $\text{--CH H}_2\text{--S}$ ); Hann J. *Chem. Soc Perkin Trans. I* 307-314 (1982) ( $\text{--CH--CH--}$ , cis and trans); Almquist et al. *J. Med. Chem.* 23:1392-1398 (1980) ( $\text{--COCH}_2\text{--}$ ); Jennings-White et al. *Tetrahedron Lett* 23:2533 (1982) ( $\text{--COCH}_2\text{--}$ ); Szelke et al. *European Appln*, EP 45665 CA (1982): 97:39405 (1982) ( $\text{--CH(OH)CH}_2\text{--}$ ); Holladay et al. *Tetrahedron. Lett* 24:4401-4404 (1983) ( $\text{--C(OH)CH}_2\text{--}$ ); and Hruby *Life Sci* 31:189-199 (1982) ( $\text{--CH}_2\text{--S--}$ ); each of which is incorporated herein by reference. A particularly preferred non-

peptide linkage is --CH<sub>2</sub>NH--. It is understood that peptide analogs can have more than one atom between the bond atoms, such as b-alanine, g-aminobutyric acid, and the like.

253. Amino acid analogs and analogs and peptide analogs often have enhanced or desirable properties, such as, more economical production, greater chemical stability, enhanced pharmacological properties (half-life, absorption, potency, efficacy, etc.), altered specificity (e.g., a broad-spectrum of biological activities), reduced antigenicity, and others.

254. D-amino acids can be used to generate more stable peptides, because D amino acids are not recognized by peptidases and such. Systematic substitution of one or more amino acids of a consensus sequence with a D-amino acid of the same type (e.g., D-lysine in place of L-lysine) can be used to generate more stable peptides. Cysteine residues can be used to cyclize or attach two or more peptides together. This can be beneficial to constrain peptides into particular conformations. (Rizo and Gierasch Ann. Rev. Biochem. 61:387 (1992), incorporated herein by reference).

## 6. Pharmaceutical carriers/Delivery of pharmaceutical products

255. As described above, the compositions can also be administered *in vivo* in a pharmaceutically acceptable carrier. By "pharmaceutically acceptable" is meant a material that is not biologically or otherwise undesirable, i.e., the material may be administered to a subject, along with the nucleic acid or vector, without causing any undesirable biological effects or interacting in a deleterious manner with any of the other components of the pharmaceutical composition in which it is contained. The carrier would naturally be selected to minimize any degradation of the active ingredient and to minimize any adverse side effects in the subject, as would be well known to one of skill in the art.

256. The compositions may be administered orally, parenterally (e.g., intravenously), by intramuscular injection, by intraperitoneal injection, transdermally, extracorporeally, topically or the like, including topical intranasal administration or administration by inhalant. As used herein, "topical intranasal administration" means delivery of the compositions into the nose and nasal passages through one or both of the nares and can comprise delivery by a spraying mechanism or droplet mechanism, or through aerosolization of the nucleic acid or vector. Administration of the compositions by inhalant can be through the nose or mouth via delivery by a spraying or droplet mechanism. Delivery can also be directly to any area of the respiratory system (e.g., lungs) via intubation. The exact amount of the compositions required will vary from subject to subject, depending on the species, age, weight and general condition of the subject, the severity of the allergic disorder being treated, the particular nucleic acid or vector

used, its mode of administration and the like. Thus, it is not possible to specify an exact amount for every composition. However, an appropriate amount can be determined by one of ordinary skill in the art using only routine experimentation given the teachings herein.

257. Parenteral administration of the composition, if used, is generally characterized by injection. Injectables can be prepared in conventional forms, either as liquid solutions or suspensions, solid forms suitable for solution or suspension in liquid prior to injection, or as emulsions. A more recently revised approach for parenteral administration involves use of a slow release or sustained release system such that a constant dosage is maintained. See, e.g., U.S. Patent No. 3,610,795, which is incorporated by reference herein.

258. The materials may be in solution, suspension (for example, incorporated into microparticles, liposomes, or cells). These may be targeted to a particular cell type via antibodies, receptors, or receptor ligands. The following references are examples of the use of this technology to target specific proteins to tumor tissue (Senter, et al., Bioconjugate Chem., 2:447-451, (1991); Bagshawe, K.D., Br. J. Cancer, 60:275-281, (1989); Bagshawe, et al., Br. J. Cancer, 58:700-703, (1988); Senter, et al., Bioconjugate Chem., 4:3-9, (1993); Battelli, et al., Cancer Immunol. Immunother., 35:421-425, (1992); Pietersz and McKenzie, Immunolog. Reviews, 129:57-80, (1992); and Roffler, et al., Biochem. Pharmacol., 42:2062-2065, (1991)). Vehicles such as "stealth" and other antibody conjugated liposomes (including lipid mediated drug targeting to colonic carcinoma), receptor mediated targeting of DNA through cell specific ligands, lymphocyte directed tumor targeting, and highly specific therapeutic retroviral targeting of murine glioma cells *in vivo*. The following references are examples of the use of this technology to target specific proteins to tumor tissue (Hughes et al., Cancer Research, 49:6214-6220, (1989); and Litzinger and Huang, Biochimica et Biophysica Acta, 1104:179-187, (1992)). In general, receptors are involved in pathways of endocytosis, either constitutive or ligand induced. These receptors cluster in clathrin-coated pits, enter the cell via clathrin-coated vesicles, pass through an acidified endosome in which the receptors are sorted, and then either recycle to the cell surface, become stored intracellularly, or are degraded in lysosomes. The internalization pathways serve a variety of functions, such as nutrient uptake, removal of activated proteins, clearance of macromolecules, opportunistic entry of viruses and toxins, dissociation and degradation of ligand, and receptor-level regulation. Many receptors follow more than one intracellular pathway, depending on the cell type, receptor concentration, type of ligand, ligand valency, and ligand concentration. Molecular and cellular mechanisms of

receptor-mediated endocytosis has been reviewed (Brown and Greene, DNA and Cell Biology 10:6, 399-409 (1991)).

**a) Pharmaceutically Acceptable Carriers**

259. The compositions, including antibodies, can be used therapeutically in  
5 combination with a pharmaceutically acceptable carrier.

260. Suitable carriers and their formulations are described in *Remington: The Science and Practice of Pharmacy* (19th ed.) ed. A.R. Gennaro, Mack Publishing Company, Easton, PA 1995. Typically, an appropriate amount of a pharmaceutically-acceptable salt is used in the formulation to render the formulation isotonic. Examples of the pharmaceutically-acceptable  
10 carrier include, but are not limited to, saline, Ringer's solution and dextrose solution. The pH of the solution is preferably from about 5 to about 8, and more preferably from about 7 to about 7.5. Further carriers include sustained release preparations such as semipermeable matrices of solid hydrophobic polymers containing the antibody, which matrices are in the form of shaped articles, e.g., films, liposomes or microparticles. It will be apparent to those persons skilled in the art that  
15 certain carriers may be more preferable depending upon, for instance, the route of administration and concentration of composition being administered.

261. Pharmaceutical carriers are known to those skilled in the art. These most typically would be standard carriers for administration of drugs to humans, including solutions such as sterile water, saline, and buffered solutions at physiological pH. The compositions can  
20 be administered intramuscularly or subcutaneously. Other compounds will be administered according to standard procedures used by those skilled in the art.

262. Pharmaceutical compositions may include carriers, thickeners, diluents, buffers, preservatives, surface active agents and the like in addition to the molecule of choice. Pharmaceutical compositions may also include one or more active ingredients such as antimicrobial  
25 agents, antiinflammatory agents, anesthetics, and the like.

263. The pharmaceutical composition may be administered in a number of ways depending on whether local or systemic treatment is desired, and on the area to be treated. Administration may be topically (including ophthalmically, vaginally, rectally, intranasally), orally, by inhalation, or parenterally, for example by intravenous drip, subcutaneous, intraperitoneal or  
30 intramuscular injection. The disclosed antibodies can be administered intravenously, intraperitoneally, intramuscularly, subcutaneously, intracavity, or transdermally.

264. Preparations for parenteral administration include sterile aqueous or non-aqueous solutions, suspensions, and emulsions. Examples of non-aqueous solvents are propylene glycol,

polyethylene glycol, vegetable oils such as olive oil, and injectable organic esters such as ethyl oleate. Aqueous carriers include water, alcoholic/aqueous solutions, emulsions or suspensions, including saline and buffered media. Parenteral vehicles include sodium chloride solution, Ringer's dextrose, dextrose and sodium chloride, lactated Ringer's, or fixed oils. Intravenous  
5 vehicles include fluid and nutrient replenishers, electrolyte replenishers (such as those based on Ringer's dextrose), and the like. Preservatives and other additives may also be present such as, for example, antimicrobials, anti-oxidants, chelating agents, and inert gases and the like.

265. Formulations for topical administration may include ointments, lotions, creams, gels, drops, suppositories, sprays, liquids and powders. Conventional pharmaceutical carriers,  
10 aqueous, powder or oily bases, thickeners and the like may be necessary or desirable.

266. Compositions for oral administration include powders or granules, suspensions or solutions in water or non-aqueous media, capsules, sachets, or tablets. Thickeners, flavorings, diluents, emulsifiers, dispersing aids or binders may be desirable..

267. Some of the compositions may potentially be administered as a pharmaceutically  
15 acceptable acid- or base- addition salt, formed by reaction with inorganic acids such as hydrochloric acid, hydrobromic acid, perchloric acid, nitric acid, thiocyanic acid, sulfuric acid, and phosphoric acid, and organic acids such as formic acid, acetic acid, propionic acid, glycolic acid, lactic acid, pyruvic acid, oxalic acid, malonic acid, succinic acid, maleic acid, and fumaric acid, or by reaction with an inorganic base such as sodium hydroxide, ammonium hydroxide,  
20 potassium hydroxide, and organic bases such as mono-, di-, trialkyl and aryl amines and substituted ethanolamines.

### **b) Therapeutic Uses**

268. Effective dosages and schedules for administering the compositions may be determined empirically, and making such determinations is within the skill in the art. The  
25 dosage ranges for the administration of the compositions are those large enough to produce the desired effect in which the symptoms disorder are effected. The dosage should not be so large as to cause adverse side effects, such as unwanted cross-reactions, anaphylactic reactions, and the like. Generally, the dosage will vary with the age, condition, sex and extent of the disease in the patient, route of administration, or whether other drugs are included in the regimen, and can be  
30 determined by one of skill in the art. The dosage can be adjusted by the individual physician in the event of any counterindications. Dosage can vary, and can be administered in one or more dose administrations daily, for one or several days. Guidance can be found in the literature for appropriate dosages for given classes of pharmaceutical products. For example, guidance in

selecting appropriate doses for antibodies can be found in the literature on therapeutic uses of antibodies, e.g., Handbook of Monoclonal Antibodies, Ferrone et al., eds., Noyes Publications, Park Ridge, N.J., (1985) ch. 22 and pp. 303-357; Smith et al., Antibodies in Human Diagnosis and Therapy, Haber et al., eds., Raven Press, New York (1977) pp. 365-389. A typical daily  
5 dosage of the antibody used alone might range from about 1 µg/kg to up to 100 mg/kg of body weight or more per day, depending on the factors mentioned above.

## 7. Chips and micro arrays

269. Disclosed are chips where at least one address is the sequences or part of the sequences set forth in any of the nucleic acid sequences disclosed herein. Also disclosed are  
10 chips where at least one address is the sequences or portion of sequences set forth in any of the peptide sequences disclosed herein.

270. Also disclosed are chips where at least one address is a variant of the sequences or part of the sequences set forth in any of the nucleic acid sequences disclosed herein. Also disclosed are chips where at least one address is a variant of the sequences or portion of  
15 sequences set forth in any of the peptide sequences disclosed herein.

## 8. Computer readable mediums

271. It is understood that the disclosed nucleic acids and proteins can be represented as a sequence consisting of the nucleotides of amino acids. There are a variety of ways to display these sequences, for example the nucleotide guanosine can be represented by G or g. Likewise  
20 the amino acid valine can be represented by Val or V. Those of skill in the art understand how to display and express any nucleic acid or protein sequence in any of the variety of ways that exist, each of which is considered herein disclosed. Specifically contemplated herein is the display of these sequences on computer readable mediums, such as, commercially available floppy disks, tapes, chips, hard drives, compact disks, and video disks, or other computer readable mediums.  
25 Also disclosed are the binary code representations of the disclosed sequences. Those of skill in the art understand what computer readable mediums. Thus, computer readable mediums on which the nucleic acids or protein sequences are recorded, stored, or saved.

272. Disclosed are computer readable mediums comprising the sequences and information regarding the sequences set forth herein. Also disclosed are computer readable  
30 mediums comprising the sequences and information regarding the sequences set forth herein wherein the sequences do not include SEQ ID Nos: 37, 38, 39, 40, 41, and 42.

## 9. Kits

273. Disclosed herein are kits that are drawn to reagents that can be used in practicing the methods disclosed herein. The kits can include any reagent or combination of reagent discussed herein or that would be understood to be required or beneficial in the practice of the disclosed methods. For example, the kits could include primers to perform the amplification reactions discussed in certain embodiments of the methods, as well as the buffers and enzymes required to use the primers as intended.

### D. Methods of making the compositions

274. The compositions disclosed herein and the compositions necessary to perform the disclosed methods can be made using any method known to those of skill in the art for that particular reagent or compound unless otherwise specifically noted.

#### 1. Nucleic acid synthesis

275. For example, the nucleic acids, such as, the oligonucleotides to be used as primers can be made using standard chemical synthesis methods or can be produced using enzymatic methods or any other known method. Such methods can range from standard enzymatic digestion followed by nucleotide fragment isolation (see for example, Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, 2nd Edition (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989) Chapters 5, 6) to purely synthetic methods, for example, by the cyanoethyl phosphoramidite method using a Milligen or Beckman System 1Plus DNA synthesizer (for example, Model 8700 automated synthesizer of Milligen-Bioscience, Burlington, MA or ABI Model 380B). Synthetic methods useful for making oligonucleotides are also described by Ikuta *et al.*, *Ann. Rev. Biochem.* **53**:323-356 (1984), (phosphotriester and phosphite-triester methods), and Narang *et al.*, *Methods Enzymol.*, **65**:610-620 (1980), (phosphotriester method). Protein nucleic acid molecules can be made using known methods such as those described by Nielsen *et al.*, *Bioconjug. Chem.* **5**:3-7 (1994).

#### 2. Peptide synthesis

276. One method of producing the disclosed proteins, such as SEQ ID NO:23, is to link two or more peptides or polypeptides together by protein chemistry techniques. For example, peptides or polypeptides can be chemically synthesized using currently available laboratory equipment using either Fmoc (9-fluorenylmethyloxycarbonyl) or Boc (*tert*-butyloxycarbonyl) chemistry. (Applied Biosystems, Inc., Foster City, CA). One skilled in the art can readily appreciate that a peptide or polypeptide corresponding to the disclosed proteins, for example, can be synthesized by standard chemical reactions. For example, a peptide or

polypeptide can be synthesized and not cleaved from its synthesis resin whereas the other fragment of a peptide or protein can be synthesized and subsequently cleaved from the resin, thereby exposing a terminal group which is functionally blocked on the other fragment. By peptide condensation reactions, these two fragments can be covalently joined via a peptide bond at their carboxyl and amino termini, respectively, to form an antibody, or fragment thereof.

(Grant GA (1992) Synthetic Peptides: A User Guide. W.H. Freeman and Co., N.Y. (1992); Bodansky M and Trost B., Ed. (1993) Principles of Peptide Synthesis. Springer-Verlag Inc., NY (which is herein incorporated by reference at least for material related to peptide synthesis).

Alternatively, the peptide or polypeptide is independently synthesized *in vivo* as described herein. Once isolated, these independent peptides or polypeptides may be linked to form a peptide or fragment thereof via similar peptide condensation reactions.

277. For example, enzymatic ligation of cloned or synthetic peptide segments allow relatively short peptide fragments to be joined to produce larger peptide fragments, polypeptides or whole protein domains (Abrahmsen L et al., Biochemistry, 30:4151 (1991)). Alternatively, native chemical ligation of synthetic peptides can be utilized to synthetically construct large peptides or polypeptides from shorter peptide fragments. This method consists of a two step chemical reaction (Dawson et al. Synthesis of Proteins by Native Chemical Ligation. Science, 266:776-779 (1994)). The first step is the chemoselective reaction of an unprotected synthetic peptide--thioester with another unprotected peptide segment containing an amino-terminal Cys residue to give a thioester-linked intermediate as the initial covalent product. Without a change in the reaction conditions, this intermediate undergoes spontaneous, rapid intramolecular reaction to form a native peptide bond at the ligation site (Baggiolini M et al. (1992) FEBS Lett. 307:97-101; Clark-Lewis I et al., J.Biol.Chem., 269:16075 (1994); Clark-Lewis I et al., Biochemistry, 30:3128 (1991); Rajarathnam K et al., Biochemistry 33:6623-30 (1994)).

278. Alternatively, unprotected peptide segments are chemically linked where the bond formed between the peptide segments as a result of the chemical ligation is an unnatural (non-peptide) bond (Schnolzer, M et al. Science, 256:221 (1992)). This technique has been used to synthesize analogs of protein domains as well as large amounts of relatively pure proteins with full biological activity (deLisle Milton RC et al., Techniques in Protein Chemistry IV. Academic Press, New York, pp. 257-267 (1992)).

### 3. Processes for making the compositions

279. Disclosed are processes for making the compositions as well as making the intermediates leading to the compositions. For example, disclosed are nucleic acids and proteins

in SEQ ID NOs:1-60. There are a variety of methods that can be used for making these compositions, such as synthetic chemical methods and standard molecular biology methods. It is understood that the methods of making these and the other disclosed compositions are specifically disclosed.

5           280. Disclosed are nucleic acid molecules produced by the process comprising linking in an operative way a nucleic acid comprising the sequence set forth herein and a sequence controlling the expression of the nucleic acid.

          281. Also disclosed are nucleic acid molecules produced by the process comprising linking in an operative way a nucleic acid molecule comprising a sequence having 80% identity  
10 to a sequence set forth in herein, and a sequence controlling the expression of the nucleic acid.

          282. Disclosed are nucleic acid molecules produced by the process comprising linking in an operative way a nucleic acid molecule comprising a sequence that hybridizes under stringent hybridization conditions to a sequence set forth herein and a sequence controlling the expression of the nucleic acid.

15           283. Disclosed are nucleic acid molecules produced by the process comprising linking in an operative way a nucleic acid molecule comprising a sequence encoding a peptide set forth in SEQ ID NO:7 and a sequence controlling an expression of the nucleic acid molecule.

          284. Disclosed are nucleic acid molecules produced by the process comprising linking in an operative way a nucleic acid molecule comprising a sequence encoding a peptide having  
20 80% identity to a peptide set forth in herein and a sequence controlling an expression of the nucleic acid molecule.

          285. Disclosed are nucleic acids produced by the process comprising linking in an operative way a nucleic acid molecule comprising a sequence encoding a peptide having 80% identity to a peptide set forth in herein, wherein any change from the herein are conservative  
25 changes and a sequence controlling an expression of the nucleic acid molecule.

          286. Disclosed are cells produced by the process of transforming the cell with any of the disclosed nucleic acids. Disclosed are cells produced by the process of transforming the cell with any of the non-naturally occurring disclosed nucleic acids.

          287. Disclosed are any of the disclosed peptides produced by the process of expressing  
30 any of the disclosed nucleic acids. Disclosed are any of the non-naturally occurring disclosed peptides produced by the process of expressing any of the disclosed nucleic acids. Disclosed are any of the disclosed peptides produced by the process of expressing any of the non-naturally disclosed nucleic acids.

288. Disclosed are animals and invertebrates produced by the process of transfecting a cell within the animal or invertebrate with any of the nucleic acid molecules disclosed herein. Disclosed are animals or invertebrates produced by the process of transfecting a cell within the animal any of the nucleic acid molecules disclosed herein, wherein the animal is a mammal  
5 invertebrate is an insect, such as drosophila. Also disclosed are animals produced by the process of transfecting a cell within the animal any of the nucleic acid molecules disclosed herein, wherein the mammal is mouse, rat, rabbit, cow, sheep, pig, or primate.

289. Also disclose are animals produced by the process of adding to the animal any of the cells disclosed herein.

## **E. Methods of using the compositions**

### **1. Methods of using the compositions as research tools**

290. The disclosed compositions can be used in a variety of ways as research tools. For example, the disclosed compositions, such as molecules disclosed herein can be used to study the interactions between the molecules, and for example, their ligands or other compounds,  
15 by for example acting as inhibitors of binding.

291. The compositions can be used for example as targets in combinatorial chemistry protocols or other screening protocols to isolate molecules that possess desired functional properties related to inhibiting DHR96 activity, for example.

292. The disclosed compositions can be used as discussed herein as either reagents in  
20 micro arrays or as reagents to probe or analyze existing microarrays. The disclosed compositions can be used in any known method for isolating or identifying single nucleotide polymorphisms. The compositions can also be used in any method for determining allelic analysis of for example, DHR96, particularly allelic analysis as it relates to xenobiotic pathway functions. The compositions can also be used in any known method of screening assays, related to chip/micro  
25 arrays. The compositions can also be used in any known way of using the computer readable embodiments of the disclosed compositions, for example, to study relatedness or to perform molecular modeling analysis related to the disclosed compositions.

## **F. Examples**

293. The following examples are put forth so as to provide those of ordinary skill in  
30 the art with a complete disclosure and description of how the compounds, compositions, articles, devices and/or methods claimed herein are made and evaluated, and are intended to be purely exemplary and are not intended to limit the disclosure. Efforts have been made to ensure accuracy with respect to numbers (e.g., amounts, temperature, etc.), but some errors and

deviations should be accounted for. Unless indicated otherwise, parts are parts by weight, temperature is in °C or is at ambient temperature, and pressure is at or near atmospheric.

**1. Example 1 The DHR96 nuclear receptor is required for xenobiotic responses in *Drosophila***

**a) Materials and Methods**

**(1) Construction of the DHR96 targeting fragment**

294. A 7.55 kb DNA fragment that contains a mutated version of the *Drosophila melanogaster* DHR96 gene was generated by introducing two deletions: (1) deleting sequences harboring the start site (26 bp) and (2) deleting the fourth exon and intron (331 bp) from the wild type sequence. In addition, a recognition site for the restriction enzyme I-Sce I was inserted into the center (cuts between position 3699 and 3700) of the 7.55 kb fragment (see fig. M1). To obtain a genomic clone DNA of the P1 clone 26-95 that harbored the complete DHR96 gene was isolated (provided by BDGP: <http://www.fruitfly.org/>). The assembly of the 7.55 kb targeting sequence was achieved by fusing three fragments:

**(a) Fragment 1 A 1.958 kb *Apa* I-Hind III fragment**

295. This was isolated by cutting P1 26-95 with Hind III and isolating a 6.599 kb Hind III fragment, which then was cut with *Apa* I and *Sgr* AI. The 1.958 kb *Apa* I – Hind III fragment was cloned into Litmus 38 (New England BioLabs) (cut with *Apa* I and Hind III).

**(b) Fragment 2 A 4.325 kb fragment**

296. This fragment contains the actual mutations and forms the core of the targeting construct. It was generated by using three pairs of PCR primers (for sequences, see oligos): (I) FAPA96 and R96EX3Sce, (II) F96Int3Sce and R96Int3, (III) F96Ex5Int3 and R96EndHind. The P1 26-95 genomic clone served as a template. Primer pair (I) produced a 1724 bp fragment, primer pair (II) a 993 bp fragment and primer pair (III) a 1650 bp fragment. The 993 bp and the 1650 bp fragments were fused in a PCR reaction using the primers F96Int3Sce and R96EndHind, generating a 2.62 kb fragment. Likewise, the 1724 bp and the 993 bp fragments were fused using the FAPA96 and R96Int3 primers to form a 2.70 kb fragment. In a final step, the 2.70 and the 2.62 kb fragments were fused using the primers FAPA96 and R96EndHind to form the aforementioned 4.325 kb fragment, which was cloned into PCR TOPO 2.1 (Invitrogen).

**(c) Fragment 3 A 1.86 kb PCR fragment**

297. Fragment 3 was generated using the primers F96Xma and R96SpeBgl, with the P1 26-95 clone as a template. The fragment was eluted and cut directly with Xma I and Spe I.

298. The 1.86 kb PCR fragment was cloned into the PCR Topo 2.1 vector (Invitrogen) containing the 4.325 kb, which was cut with Xma I and Spe I. The resulting clone was cut with Apa I and Spe I and fused to the 1.958 kb fragment, which had been previously isolated from Litmus 38 (New England Biolabs) with Apa I and Spe I. The resulting clone is the 7.55 kb targeting fragment. A sequence printout and annotation of this fragment is included (SEQ ID NO:37).

### **(2) Construction of the hs-Gal4-DHR96 fusion gene**

299. A fusion of the Gal4 DNA binding domain (amino acids 1 to 147) and the DHR96 hinge region and ligand binding domain (LBD) (amino acids 99 to 723) was generated to create a Gal4-LBD fusion protein. Two PCR fragments were generated: (I) a 475 bp fragment using the primers FGALXB and RGAL96 and a Gal4 containing plasmid as a template. (II) F96BEG and R96/936 generate a 372 bp fragment from pLF20N, which contains the DHR96 cDNA (Fisk and Thummel, 1995). Fragments (I) and (II) possess a 15 bp overlap that was then utilized to fuse them by PCR. The resulting 832 bp fragment was cut with Xba I and Age I and cloned into pLF20N, which had been cut with the same enzymes to remove the DHR96 DNA-binding domain. The resulting plasmid is termed pGAL96. To obtain the final transformation vector, the Gal4-DHR96 fusion gene was isolated from pGAL96 with Not I and Nhe I and ligated to pCASPER hs-act cut with Xba I and Not I (SEQ ID NO:38, (see Seq 2 for the sequence of the insert in this vector, encoding the Gal4-LBD fusion).

### **(3) Construction of the hs-DHR96 RNAi vector**

300. An inverted repeat sequence that corresponds to a part of the coding region for the DHR96 ligand-binding domain (each repeat corresponds to nucleotides 1444-2371 of the DHR96 plasmid pLF20N; Fisk and Thummel, 1995) was generated. The repeats are separated by a unique spacer region of 101 bp that corresponds to nucleotides 2372-2472 of the same DHR96 cDNA. Two primer pairs were used: (I) F96Xbai and R96BspE1 and (II) F96Xbai and R96BspE2. Both fragments were cut with Bsp EI and ligated. The ligated fragment was purified and cut with Xba I and cloned into Litmus 28 (New England Biolabs) cut with Xba I. After the cloned fragment (1956 bp) was verified by restriction analysis, it was excised with Xba I and inserted into pCasper hs-act cut with Xba I.

### **(4) Construction of the hs-DHR96 vector and fly transformation**

301. This vector produces wild type DHR96 protein under the control of an hsp70 promoter in a transgenic animal. A full length cDNA was excised from the plasmid pLF20N

with the restriction enzymes Not I and NheI and cloned it into pCasper hs-act vector cut with Not I and Xba I. Transformant flies were isolated using standard methods (Rubin GM, Spradling AC. Genetic transformation of *Drosophila* with transposable element vectors. *Science*. 1982 Oct 22;218(4570):348-53).

#### (5) Construction of pET24c-DHR96

302. To generate antibodies, DHR96 antigen was produced from a 1.8 kb EcoRV fragment (597 amino acids), which includes most of the cDNA, but excludes the DNA binding domain. The 1.8 kb Eco RV fragment was isolated from pLF20, a plasmid that contains a full length DHR96 cDNA (pLF20 differs from pLF20N in the following: pLF20 was cut with HindIII, filled in, and religated to create a unique Nhe I site. The new plasmid was termed pLF20N). pET24c (Novagen) was cut with Bam HI and Xho I and blunt ends were generated by fill-in, and subsequently the Eco RV fragment was cloned into this vector. Orientation was tested using restriction analysis. A sequence printout of this clone is included (SEQ ID NO:39Seq. 3).

#### (6) Construction of pMAL-DHR96

303. To purify antisera, soluble DHR96 protein was produced by fusing the original antigen to the Maltose-binding protein. To subclone the Eco RV fragment of DHR96 (the original antigen coding section) into pMAL-c2X (New England Biolab), a fragment from pET24c-DHR96 was PCR amplified by using the primer pair F96ANhe and R96AHind. The fragment was cut directly with Nhe I and HindIII and cloned into pMAL-c2X cut with Xba I and HindIII.

#### (7) Oligonucleotides

##### Oligonucleotides

SEQ ID NO:40	F96Xma	5'-GAGAGATGTGCTTCGTAAAGCATCAACCC
SEQ ID NO:41	R96SpeBgl	5'-GGACTAGTAGATCTAGAGGATTCTACAAATGTCCAGTGTCTCCC
SEQ ID NO:42	R96Int3	5'-CCATTATTATCGCCATAATCGTAAAGG
SEQ ID NO:43	R96EX3SCE	5'-ATTACCCTGTTATCCCTAGCGGGTTACCTTAATGCGATCATCGCCC
SEQ ID NO:44	R96endhind	5'-GGAAAGCTTTTCCTGCTGATCAATAATACC
SEQ ID NO:45	FAPA96	5'-TGGGCCCATCACTTGCTTGTAACCGCCGAAGAACTGCGCGG
SEQ ID NO:46	F96INT3SCE	5'-CGCTAGGGATAACAGGGTAATAACAGTCCACGGTATTAGCCTATAGG
SEQ ID NO:47	F96EX5Int3	5'-CGATTATGGCGATAATAATGGCCAAAGAGAACATGGGCAACATACGC
SEQ ID NO:48	FGALXB	5'-GAAGCAAGCCTCTAGAAAGATGAAGC
SEQ ID NO:49	RGAL96	5'-CGTGCCGTTCTCCATCGATACAGTCAACTGTCTTTGACC

SEQ ID NO:50	R96/936	5'-GCCTGGATAGTCGATCAAATGCG
SEQ ID NO:51	F96BEG	5'-ATGGAGAACGGCACGGATGC
SEQ ID NO:52	F96XBAi	5'-TACATTCTAGAGACCAACTACAACGACGAGCCCAGTCTGG
SEQ ID NO:53	R96BspE1	5'-CATTCATCCGGACATTAATTATGAACTTGTTTCAGACGCTCC
SEQ ID NO:54	R96BspE2	5'-GGGCATCAACTCCGGAATTAAATGCCCCGACACGCATCGG
SEQ ID NO:55	RPAXCRE-AN	5'-GTCTCACGACGTTTTGAACCCAGAAATCGAGCTCGCCCCGGGG
SEQ ID NO:56	RPAXCRECO	5'-CACGAATTCCAAACTGTCTCACGACGTTTTGAACCC
SEQ ID NO:57	FPAXFSE-AN	5'-GAGAGCTAGCATGCCGGCTAGATCTCGAGATCGGCCGGCCTAGG
SEQ ID NO:58	FPAXPOLY	5'-GAACTGCAGCTCGAGAGCTAGCATGCCGGC
SEQ ID NO:59	F96ANhe	5'-GGAGATATACATATGGCTAGCATGACTGGTGG
SEQ ID NO:60	R96AHind	5'-TGCTCGAAGCTTCGCAGAAGATAATAGTAGG

#### (8) DHR96 gene targeting

304. The 7.55 kb genomic fragment containing a mutated DHR96 gene (see above) was inserted into the *Drosophila* genome as described (Rong YS, Golic KG. Gene targeting by homologous recombination in *Drosophila*. Science. 2000 Jun 16;288(5473):2013-8). w; [hsp70-FLP]4 [hsp70 I Sce I]2b Sco/S2 CyO females were crossed to w; [<(96TG GFP+> w+)] males that carried the targeting fragment on the second chromosome. Larvae were heat shocked during the third larval instar to trigger targeting events in the germline of females. [hsp70-FLP]4 [hsp70 I Sce I]2b Sco/ [<(96TG GFP+> w+)] females were then collected and crossed them to w; Ser1/TM6B, Tb males. 918 vials of such crosses (5 males and 10 females) were set up which generated approximately 150,000 flies that were screened for GFP+, but white-eyed individuals. These flies were crossed to w1118; Ly/TM6C Tb Sb, and stocks were subsequently established from a single chromosome. The DHR96E25 allele was isolated from one of these stocks.

#### (9) Reduction of the DHR96 targeted event to a single copy by I-CreI

305. Males carrying the tandem duplication allele (w1118/Y; DHR96E25/DHR96E25) were mated to v hsp70 CreI; Sb/TM6 females in mass. After 3 days at 25°C, the parental flies were removed and the progeny were heat-treated at 36°C for one hour to induce CreI recombinase. Males that eclosed were individually mated to w1118; Ly/TM6C females. One male progeny (w1118/Y; DHR96Cre reduced/TM6C) that had lost GFP expression (indicating a recombination event had occurred) was selected from each vial and individually mated to

w1118; Ly/TM6C females to establish a stock containing the reduced allele (Rong and Golic 2002). Mutant strains were characterized by Southern blotting, PCR, and DNA sequencing using standard methods. The DHR9616A mutant stock was selected for further characterization.

#### (10) Tissue antibody stains

5           306. Wandering third instar larval tissues were dissected and fixed as previously described (Boyd, L., O'Toole, E. and Thummel, C.S. (1991). Patterns of E74A RNA and protein expression at the onset of metamorphosis in *Drosophila*. *Development* 112, 981-995). DHR96 protein was detected with anti-DHR96 antibodies diluted 1:100 and incubated overnight at 4 °C. Donkey anti-rabbit CY3 secondary antibodies (Jackson) were used at a 1:200 dilution as a  
10 secondary antibody. The stains were visualized on a Biorad confocal laser scanning microscope.

#### (11) Western blots analysis

          307. Protein from adult flies was extracted by grinding flies in SDS sample buffer and boiling. The equivalent of approximately one adult fly was loaded in each lane of an 8% polyacrylamide gel, separated by electrophoresis and transferred to PVDF membrane.  
15 Ectopically expressed DHR96 protein was produced by heat-treating flies at 37.5 °C for 30 minutes followed by a three hour recovery at room temperature before the extraction procedure. DHR96 protein was detected by incubating the membrane first with a 1:500 dilution of anti-DHR96 affinity purified antibodies followed by a 1:1000 dilution of goat anti-rabbit HRP secondary antibody (Pierce). A supersignal chemiluminescence kit was used to develop the  
20 signal (Pierce).

#### (12) Toxicity assays

          308. Adult flies were raised on standard cornmeal/agar food and starved overnight under humid conditions at 25 °C before treatment with DDT. A DDT stock solution was prepared by dissolving crystalline DDT (Sigma) in 100% ethanol. Appropriate DDT dilutions  
25 were made by diluting the DDT stock with 5% sucrose and pipetting 275 µl of the solution onto a strip of Whatman filter paper inside a small glass scintillation vial. Twenty adult flies were placed in each vial which was plugged with cotton. Mortality was scored 10 hours later at room temperature. For each DDT concentration, three replicates, each of twenty adult flies, were used. For the time course assay, 100 ng/µl of DDT was used and mortality scored every hour for 10  
30 hours.

**b) Results****(1) DHR96 is closely related to known xenobiotic receptors**

309. The phylogenetic relationship of DHR96 to other nuclear receptors was investigated for information related to function. When performing a BLASTP search, the closest  
5 homolog to DHR96 in vertebrates is the Vitamin D3 Receptor (VDR). The Pregnane X Receptor (PXR) as well as the Constitutively Androstane Receptor (CAR) comprise other high scoring homologs. (Fig. 1).

**(2) DHR96 is expressed in the alimentary canal, the salivary glands and the fat body**

10 310. Antibody stains of third instar larvae were used to analyze whether DHR96 would be expressed in tissues that function in detoxification. DHR96 antibodies strongly stain tissues of the alimentary canal (Fig. 2). In particular, the gastric caeca, the major site of absorption in Diptera, show a much stronger staining than the remainder of the midgut, which also plays a role in nutrient absorption. Strong expression in the Malpighian tubules, the principal excretory  
15 organ in insects, was also observed. The excretory system maintains homeostasis, controlling salt levels and osmotic pressure, but is primarily responsible for the removal of harmful metabolites such as nitrogenous wastes derived from purine metabolism, or toxic compounds that were absorbed from the food. Outside the alimentary canal, strong staining in the salivary gland and the fat body were detected. The insect fat body is the functional equivalent of the  
20 mammalian liver, because it is the principal site of intermediary metabolism and detoxification. Taken together, the finding that DHR96 expression is tightly associated with tissues known to be involved in detoxification provides strong support for the proposal that DHR96 functions in a xenobiotic pathway.

**(3) DHR96 function is dispensable under standard conditions**

25 311. RNA interference (RNAi) and gene targeting were used to disrupt *DHR96* function because no existing mutants were available. The effects of *DHR96* RNAi were analyzed by generating transgenic lines that express snapback RNA under the control of a heat-inducible promoter. Three independent lines showed strong reduction of *DHR96* mRNA in northern blots when treated with a single heat-shock, but displayed no discernable phenotype.  
30 Using a variety of heat-shock regimens, e.g. longer single and double treatments or 12 hr repetitions, did not affect the outcome of this observation. These findings suggest that *DHR96* mRNA is not necessary for viability under standard conditions, indicating either that DHR96 protein is very stable or dispensable for survival.

312. Gene targeting (Rong, Y. S., and Golic, K. G. (2000). *Science* 288, 2013-2018) was used to generate mutations in *DHR96* because no deficiencies or P elements were known in this region of the genome. As a first step, the gene targeting procedure requires classical P-element transformation in order to generate transgenes that harbor the targeting sequence flanked by *FRT* sites. The targeting DNA is then mobilized and turned into a linear, recombinogenic molecule *in vivo* by activating the *FLP* recombinase and the endonuclease *I Sce I*. As a consequence of this targeting technique, which is based on an “ends-in” mechanism, the resulting mutation is basically a replacement of the original gene with a tandem duplication of two mutant copies (Fig. 3). Mutations were engineered in such a way that both copies would result in non-functional gene products. In particular, a region around the translation start site (25 bp), and the complete sequence of exon four was deleted, the downstream intron, and the splice acceptor site at exon 5 (together ~300 bp). These mutations should lead to a block in translation initiation as well as removal of most of the ligand binding domain of the receptor. We constructed a targeting vector that contained two eye markers: *pax6-EGFP* and *mini-white*. Once mobilized by the *FLP* recombinase, the *EGFP* gene separates physically from the *mini-white* gene, which lies outside the *FRT* sites. Consequently, the subsequent strategy employed to identify potential targeting events is based on the presence of the *EGFP* marker and the simultaneous absence of the *mini-white* marker in the eye.

313. In a screen of ~150,000 flies, a total of 42 events were detected. Of these, 18 mapped to the third chromosome, which harbors the *DHR96* gene. At least one of the 18 events was identified as a targeting event in the *DHR96* gene, and we termed this allele *DHR96<sup>E25</sup>*. To avoid problems that might arise from the truncated protein in the *DHR96<sup>E25</sup>* mutant, we decided to reduce the existing duplication to one mutant copy by utilizing the *I Cre I* site that was built into the targeting vector, essentially following the procedure described by (Rong, Y. et al., (2002) *Genes Dev* 16, 1568-1581). This procedure yielded a new *DHR96* allele, *DHR96<sup>16A</sup>*, which, based on sequence and western analysis, constitutes a protein null. Several lines of evidence suggest that these alleles represent specific targeting events in the *DHR96* gene. First, genomic Southern blots of animals homozygous for the targeting events displayed the predicted fragment patterns of a tandem duplication (*DHR96<sup>E25</sup>*) or a reduced single copy (*DHR96<sup>16A</sup>*). Second, northern analysis revealed the absence of the wild type mRNA in the mutant animals. Third, antibody stains and Western analysis show a strong reduction or absence of the *DHR96* protein in *DHR96<sup>16A</sup>* or *DHR96<sup>E25</sup>* flies (add fig for this). Fourth, Southern blot hybridization and

sequencing of PCR products demonstrated that exon/intron 4 of wild type *DHR96* is absent in homozygous *DHR96*<sup>16A</sup> or *DHR96*<sup>E25</sup> animals.

314. Flies homozygous for *DHR96*<sup>E25</sup> or *DHR96*<sup>16A</sup> are viable and fertile when grown on standard cornmeal food. However, when placed on instant food (Carolina 424) in the absence  
5 of yeast, viability decreases to about 1%, whereas wild type flies do comparably well with a survival rate of ~35% compared to standard food. Interestingly, the addition of yeast restores viability to 100%. This suggests that either *DHR96* is required for the proper execution of certain nutritional pathways, or that *DHR96*<sup>E25</sup> larvae fail to neutralize toxic metabolites that are produced when animals are reared on nutritionally poor media. To test the possibility that  
10 *DHR96* mutants have a decreased tolerance for toxins, it was determined whether *DHR96* is expressed in tissues that are known to play critical roles in the detoxification process.

#### **(4) *DHR96* mutants display reduced viability in the presence of DDT**

315. As a test of *DHR96* acting in a xenobiotic pathway, *DHR96* mutants were tested  
15 for sensitivity to the pesticide DDT. Adult wild type flies (Canton S) and *DHR96*<sup>16A</sup> were exposed or *DHR96*<sup>E25</sup> flies to varying concentrations of DDT and recorded survival rates after a fixed time. The findings showed that *DHR96* mutants were more sensitive to DDT and died at lower concentrations of DDT compared to control animals (Fig. 4A). In addition, when challenged with a fixed concentration of DDT, *DHR96* homozygotes died more rapidly than wild  
20 type flies (Fig. 4B). Taken together, these results indicated that *DHR96* is required for natural resistance levels to the pesticide DDT, and that *DHR96* functions in a xenobiotic response pathway.

316. In addition to DDT, the outcrossed lines were tested for sensitivity to phenobarbital (a well characterized cytochrome P450 agonist), and tebufenozide (an insect  
25 growth regulator that is widely used in agricultural applications). The adult Canton S flies and the *DHR96*<sup>E25</sup> outcrossed lines were exposed to varying concentrations of drug and recorded effects after a fixed time (Fig. 11). DDT was assayed by starving young healthy adult flies overnight and then transferring them to vials, in three groups of 20 flies each, with filter paper soaked with 5% sucrose alone or 5% sucrose and DDT at different concentrations. The number  
30 of living flies was scored after 23 hours. Phenobarbital was tested in the same way, except that the number of actively moving flies was scored after 23 hours. Tebufenozide was administered to larvae in the food, and the number of surviving adult flies was scored. These studies showed that, whereas the original *DHR96*<sup>E25</sup> mutant line is more sensitive than Canton S to DDT

treatment, this sensitivity must be due to a difference in genetic background since the outcrossed line showed no such sensitivity to this compound (Fig. 11A). In contrast, both the original and outcrossed DHR96E25 mutant lines are more sensitive to phenobarbital than Canton S, indicating that the genetic background did not contribute to this effect (Fig. 11B). Treatment with tebufenozide resulted in a slight sensitivity of the outcrossed DHR96E25 mutant to this compound (Fig. 11C). Taken together, these results indicate that DHR96 is required for natural resistance levels, showing it acts in a xenobiotic response pathway.

#### (5) Overexpression of DHR96 has no effect on viability

317. Most nuclear receptors cause lethality when overexpressed, indicating that these proteins do not require an obligatory ligand for some or even all of their functions. To analyze whether DHR96 would disrupt essential pathways and cause lethality when expressed ectopically, a transgenic line that harbored a full-length *DHR96* cDNA under the control of a heat-inducible promoter was produced. Western and Northern analysis showed that heat-treated larvae and flies carrying this construct generated at least 100 times more *DHR96* mRNA and protein than wild type flies lacking the transgene. Nevertheless, overexpression of this protein did not result in any visible effect, suggesting two possible scenarios: (I) DHR96 activity requires binding to a ligand or a protein partner, or (II) DHR96 target genes do not function in vital pathways, at least not under standard laboratory conditions. Naturally, both possibilities may be true. Microarray experiments were used to dissect how DHR96 might function on the molecular level.

#### c) Microarray experiments

318. As a first step toward identifying target genes regulated by DHR96, the protein was overexpressed in larvae and analyzed its effects on gene expression by microarray analyzed. Affymetrix oligonucleotide chips designed to detect ~13,200 genes (the majority in the fly genome) were used, the raw data with dCHIP (Li C, Wong WH. Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. Proc Natl Acad Sci U S A. 2001 Jan 2;98(1):31-6; Li, C., and Wong, W. H. (2001) Genome Biol 2, 0032.1-0032.11; <http://www.dchip.org/>) was analyzed, and filtering with Microsoft Access was performed. After rigorous filtering, only 72 genes remained that had a higher than 1.8-fold change when compared to the controls. Interestingly, of the top 20 reduced genes, six are members of all four major detoxification gene families, which comprise a total of 198 members in *Drosophila*. This represents a highly significant result ( $p=2.8 \times 10^{-27}$ , based on  $\chi^2$ ), because the chances of picking 6 of these genes in a random sample of 20 genes are more than 20-fold lower

than the observed number. Interestingly, no such concentration of genes encoding detoxifying enzymes exists on the list of induced genes, suggesting that DHR96 may repress these genes in the absence of suitable ligands.

319. Further examination of this list reveals other genes that can contribute to a  
5 xenobiotic response pathway. The top down-regulated gene (25-fold by dChip) encodes Lsp1-g, which is synthesized by the fat body and constitutes one of the most abundant proteins in the insect hemolymph. This protein is thought to act as a storage reservoir for nutrients during metamorphosis although it has also been proposed to transport small hydrophobic compounds within the circulatory system. The remaining down-regulated genes include three cuticle genes  
10 and one gene involved in cuticle tanning (black), consistent with the known role for cuticle deposition in toxin defense (Wilson et al. *Ann. Rev. Entomol.* 46:545-71, 2001). Other genes include a disproportionately large number that encode enzymes, such as a carboxylesterase, seven serine proteases, ornithine decarboxylase-1, dopamine N-acetyltransferase, an oxidoreductase, a g-butyrobetaine dioxygenase, a putative glucosidase, a chitin binding protein,  
15 and a transporter. Many genes that are up-regulated upon ectopic DHR96 expression) also have functions consistent with detoxification, including two cytochrome P450 genes (Cyp4p1, Cyp12d1-d). Only four families of cytochrome P450s are known to play a role in pesticide resistance: Cyp4, Cyp6, Cyp9, and Cyp12, each of which are represented in our microarray results (Ranson et al. *Science*, 298:179-81, 2002; Hemingway et al. *Insect Biochem Mol Biol*,  
20 34:653-65, 2004). A range of enzyme-encoding genes were also detected, including the neuralized ubiquitin-protein ligase gene, phr DNA repair enzyme, eTrypsin, mitochondrial carnitine palmitoyltransferase I, a phosphatidate phosphatase gene (wunen-2), a oxidoreductase-encoding gene, a lysosomal transport gene, the drosomycin-2 defense response gene, a glycine dehydrogenase gene, two genes encoding chitin binding proteins (CG10140, CG7714), and,  
25 interestingly, SCAP, which encodes the fly ortholog of the mammalian protein that releases sterol regulatory element binding-protein (SREBP) from intracellular membranes in response to sterol depletion. This set of 72 DHR96-regulated genes appears to represent a coordinated genomic response to xenobiotics.

## 2. Example 2

### 30 a) GAL4-DHR96/LBD experiments

320. To determine if DHR96 is activated by the pesticide DDT the methods disclosed herein can be used. Flies containing two different transgenes will be mated together allowing us to directly assay for DHR96 LBD activation in vivo (for detailed methods and description of

vectors see: (Kozlova, T., and C.S. Thummel (2003) Methods to characterize *Drosophila* nuclear receptor activation and function in vivo. In: "Methods in Enzymology. Nuclear Receptors, Vol. 364 (Russell, D.W., and Mangelsdorf, D.J., eds.), Academic Press, New York, pp. 475-490.)). One transgene is under the control of a heat-inducible promoter and contains the GAL4 DNA binding domain fused to the DHR96 ligand binding domain. The second transgene contains a GAL4-dependent GFP or lacZ reporter gene (Kozlova, T., and C.S. Thummel (2003) Methods to characterize *Drosophila* nuclear receptor activation and function in vivo. In: "Methods in Enzymology. Nuclear Receptors, Vol. 364 (Russell, D.W., and Mangelsdorf, D.J., eds.), Academic Press, New York, pp. 475-490.)). Upon heat induction, GAL4-DHR96 LBD protein can bind to the UAS-GFP or UAS-lacZ reporter. In the absence of a ligand, the reporter will not be activated; however, in the presence of a ligand, the GAL4 DHR96 LBD protein can be switched into an active conformation and induce reporter gene expression (Kozlova, T., and C.S. Thummel (2003) Methods to characterize *Drosophila* nuclear receptor activation and function in vivo. In: "Methods in Enzymology. Nuclear Receptors, Vol. 364 (Russell, D.W., and Mangelsdorf, D.J., eds.), Academic Press, New York, pp. 475-490.); Kozlova, T. and Thummel, C.S. (2002). Spatial patterns of ecdysteroid receptor activation during the onset of *Drosophila* metamorphosis. *Development* 129, 1739-1750).

321. To determine if drugs, such as DDT, can activate the DHR96 GAL4-LBD construct, two developmental stages will be tested. First, organs from late third instar larvae that have both transgenes will be dissected and cultured in the presence of several different concentrations of drug and assayed for reporter gene expression. Second, if activation of the GAL4-LBD construct by drug requires either ingestion of the toxin or contact with the cuticle of the fly, adults will be heat-shocked to induce the GAL4-LBD construct, placed in scintillation vials containing drug, as previously above in the toxicity assays, and assayed for induction of reporter gene expression in adult tissues. Changes in the activity of the reporter gene in the presence, but not the absence, of drug will be an indication that that compound is having a direct effect on the activity state of the DHR96 LBD.

322. Disclosed are systems that can identify ligands, such as hormones, for nuclear receptors, such as *drosophila* nuclear receptors. There are many members of the nuclear receptor superfamily for which there is no known ligand — the so called orphan nuclear receptors. It is desirable to link these receptors to a ligand if it exists.

323. One way of identifying ligands for nuclear receptors involves expressing a fusion of the GAL4 DNA binding domain to a nuclear receptor ligand binding domain (LBD), in

combination with a GAL4-responsive reporter gene. The fusion protein is inactive unless its hormone is present, allowing it to switch into an active conformation and turn on the GAL4-responsive reporter, such as a lacZ report giving a color readout. In one variation of this method, which has been widely exploited by pharma companies for high throughput screens, stably  
5 transfected tissue culture cells of different cell types are used for the cell background to perform the assay. One way to do this assay would be use every tissue in the animal as a context for screening for hormones, not just a tissue culture cell where the appropriate cofactors or partner transcription factors might be missing, because presumably every cell has a different molecular background.

10 324. One method used to get around this problem in mice is disclosed in WO 00/17334 for "Analysis of ligand activated nuclear receptors (in vivo)" by Solomon et al. (See also, Solomin, L., et al., (1998). Nature 395, 398-402). This system was designed for the mouse, because the GAL4 system of linking the GAL4 DBD to a particular LBD works poorly in mouse.

15 325. Disclosed herein is a system for drosophila for identifying ligands for nuclear receptors, where the GAL4 system works very well for driving tissue- and stage-specific ectopic gene expression. The system typically utilizes a heat-inducible promoter to widely express the GAL4-LBD fusion proteins, but any inducible promoter can be used. This allows monitoring of activation in all tissues both spatially and temporally. The pattern of lacZ expression in animals so transformed allows visualization of where and when a particular LBD is active during  
20 development, guiding one towards possible sources of hormone.

326. This has been used to show the patterns of GAL4-EcR and GAL4-USP activation during the onset of metamorphosis accurately reflect what would be expected for regulation of EcR/USP by its hormone, 20-hydroxyecdysone (Kozlova, T. and Thummel, C.S. (2002). Spatial patterns of ecdysteroid receptor activation during the onset of Drosophila metamorphosis.  
25 Development 129, 1739-1750). Spatial patterns of ecdysteroid receptor activation during the onset of Drosophila metamorphosis. Development 129, 1739-1750). This system has also been used to show that an orphan nuclear receptor, DHR38, is activated by a unique set of ecdysteroids in the animal (Baker, K. D., et al., (2003). The Drosophila orphan nuclear receptor DHR38 mediates an atypical ecdysteroid signaling pathway. Cell 113, 731-742).

30 327. Disclosed herein are hsp70-GAL4-LBD transformants for all 18 Drosophila nuclear receptors. The activation patterns of these constructs have been characterized during embryogenesis and the onset of metamorphosis. These constructs can be used with a UAS-GFP reporter to simplify the readout of activation, paving the way for compound screens.

328. These constructs can be used to screen compounds for ligand activity. For example, a collection of pesticides can be found in the Agro plate (see <http://www.msdiscovery.com>). Other plates can also be found at Micro Source Discovery, and are herein incorporated by reference at least for compound libraries and their contents. They also list plates of available collections of natural compounds.

### 3. Example 3: Effective assays for studying drug sensitivity in *DHR96* mutants.

329. Two contact poisons, DDT and tebufenozide, as well as the GABA agonist, Phenobarbital, have been tested. This set of compounds can be expanded to include the major classes of pesticides used for insect control, all of which have been compromised to some extent by adaptive resistance in pest species. These major classes include organochlorines, organophosphates, carbamates, pyrethroids, nicotinoids, and insect growth regulators. Representative compounds from these classes are shown in Table 3, along with their solubility. They include several compounds that have been used in studies of *C. elegans* and vertebrate xenobiotic responses, as well as paraquat to test responses to oxidative stress. Methyl parathion can also be tested, which is a weak insecticide, but which becomes a potent acetylcholinesterase inhibitor (methyl paraoxon) upon metabolism. *DHR96* mutants can be less sensitive to this compound than wild type. Imidacloprid, a nicotinoid that that is one of the most widely used insecticides worldwide, fipronil which has both pet and agricultural applications and acts as a GABA antagonist, or additional pyrethroids can also be tested.

**Table 4. List of compounds:**

Compound	Description	Solubility
DDT	Organochlorine, contact poison, thought to target sodium channels	ethanol
Phenobarbital	GABA mimetic, causes paralysis	water
Permethrin	Pyrethroid, blocks voltage gated sodium channels	comes as liquid
Sodium diethyldithiocarbamate trihydrate	Carbamate, cholinesterase inhibitor	water
Carbaryl	Carbamate, cholinesterase inhibitor	water
Methyl parathion	Organophosphate, contact poison	acetone
Malathion	Organophosphate, contact poison	comes as liquid
Propetamphos	Organophosphate contact poison, cholinesterase inhibitor	comes as liquid
Tebufenozide	Contact poison, ecdysone agonist	ethanol
Nicotine	Contact poison	water
Nithiazine	Neonicotinoid, used on plant sucking insects	water
Methoprene	JH mimetic, insect growth regulator	ethanol
PCN	Synthetic hormone that induces P450s in vertebrates	DMSO
Rifampicin	Antibiotic that inhibits RNA polymerase, used in vertebrate xenobiotic studies	DMSO
Colchicine	Alkaloid that inhibits mitosis, used in vertebrate xenobiotic studies	ethanol
Paraquat	Generates oxygen radicals, inducing stress and decreasing life span, induces GSTs which can provide resistance to oxidative stress	water

330. The key to defining the sensitivity of *DHR96* mutants to toxic compounds is the development of effective and reproducible assays for drug delivery. To feed compounds to adult insects, the method for administering the mutagen ethylmethane sulfonate (EMS) (Lewis et al. Dros Info. Serv. 43:193, 1968) can be used. Young adult flies, within the first five days of their life, are starved overnight in an empty vial and then transferred to a vial that contains 5% sucrose and different concentrations of the drug to be tested. The flies congregate on the filter paper to drink the sugar solution along with the drug. This method of application also provides significant surface contact as well as possible fumigant modes of entry through the tracheal system. This assay has not resulted in detectable differences in the behavior of wild type and *DHR96* mutant flies, indicating that there are no obvious differences in taste reception, or eating and drinking behavior that might result in different doses of drug between mutant and control. For all of our drug treatment studies, the highest concentration of vehicle alone is tested to determine that it does not have an effect on the experiment. An initial dose-response curve using 10-fold changes in drug concentration for either 10 or 24 hours can be used. Treatment with each drug concentration is performed in triplicate, with 20 adult flies per vial. These numbers can be increased as well, although this has not had a significant effect on experimental variability in past studies. These initial dose-response curves result in the identification of a concentration at which most animals survive as well as a higher concentration that kills most animals. The study is then repeated using 2- to 3-fold differences in dose spanning this critical range of concentrations. This provides us with a lethality curve, error bars for each data point, and an LD50 that can be compared between mutant and wild type. If desired, a time course study at a fixed concentration of pesticide can also be conducted using a similar assay.

331. A method used in other insects to assay contact toxins in *Drosophila* can also be used (Daborn et al. Mol Genet Genomics, 266:556-63, 2001). Different amounts of the compound to be tested are mixed with 200  $\mu$ l acetone and added to a glass scintillation vial. The vial is rolled so that the liquid contacts all glass surfaces. This is continued until the acetone has evaporated, leaving the toxin evenly distributed inside the vial. Groups of 20 young adult flies are transferred to each vial and lethality is scored after a fixed time. Alternatively, a fixed compound concentration is tested over a range of times. The determination of appropriate doses and treatment times is similar to that described above for the adult feeding assay. This method has been used successfully in to generate a lethality curve for Canton S wild type animals treated with DDT.

332. The above assays are for adult toxicity studies, scoring the number of dead flies resulting from exposure. Not all compounds, however, result in lethality. For example, phenobarbital increases the chloride current from the GABA receptor, enhancing the effects of this inhibitory neurotransmitter (Barber et al., Proc R Soc Lond B Biol Sci 206:319-27, 1979).

5 This compound is used clinically in humans as an anticonvulsant. At high doses in insects, it results in ataxia and, eventually, lethality. The experiment depicted in Figure 11B shows that *DHR96* mutants display a significant sensitivity to this compound relative to the Canton S control, a result we have seen reproducibly. Standardized assays have been developed to characterize behavioral defects in *Drosophila* (Bainton et al., Curr Biol 10:187-94, 2000; Rival  
10 et al. Curr Biol 14:599-605, 2004). Several of these can be employed to quantitate the effects of phenobarbital and similar drugs that result in abnormal behavior. First, running ability can be tested by transferring eight young adult flies, either *DHR96* mutants or Canton S control, into a 10 ml plastic pipette. Both ends are sealed with parafilm and one half of the pipette will is inserted into a hole in a black foam block such that the pipette is held horizontally, allowing the  
15 flies to run along its length. A fiber optic lamp is placed at the opposite end of the pipette to create a clear gradient from dark to light, to stimulate a phototactic response. For each test, the flies are knocked into the dark half of the pipette and then returned to the horizontal test position. The time is recorded at which the first six flies enter the light half of the pipette. Four trials will be done for each set of eight adults tested. The resulting times are used to calculate mean  
20 performance coefficients, as described (Palladino et al. Genetics 161:1197-208, 2002). Statistical analysis of the data can be performed using a Student's *t*-test.

333. The second behavioral assay is a flight ability assay, performed essentially as described (Benzer et al. Sci Am 229:24-37, 1973). Twenty young adult mutant or wild type flies are dumped into a glass funnel placed on top of a 500 ml graduated cylinder, such that they are  
25 released into the cylinder near the 500 ml mark on top. The glass cylinder is coated with paraffin oil to provide a sticky surface to which flies will adhere. Healthy animals initiate flight immediately and thus tend to become caught near the opening of the funnel. Weaker flying animals, in contrast, fall farther toward the bottom before being caught. Performance coefficients are calculated for the population added to the cylinder by assigning a numerical  
30 score for the distance fallen by each fly, as described (Palladino et al). Statistical analysis of the data can be performed using a Student's *t*-test.

334. Finally, the most widely used behavioral assay for measuring locomotor activity, called a climbing assay or negative geotaxis assay is used. Twenty young adult flies are placed

in a 250 ml graduated cylinder and the top is sealed with parafilm. The flies are knocked gently to the bottom of the cylinder and then allowed to climb for one minute. The number of flies in the top, middle, or bottom one-third is determined and recorded. This can be further subdivided if necessary. Three trials are performed with one population of flies, and the results are

5 averaged. The mean number of flies in each region of the cylinder can be calculated as a fraction of the total population of flies, and a performance index is determined as described (Rival et al.). Statistical analysis of the data will be performed using a Student's *t*-test. A more general motility assay can also be used in which flies are treated with drug and then transferred to a regular vial without food. The flies are gently banged into the bottom of the vial, the top is removed from  
10 the vial, and the flies are allowed to escape for a fixed period of time before the top is resealed. The number of remaining flies is then scored and an average is calculated from several repeated tests of the same population.

335. An advantage to non-lethal drugs such as phenobarbital is that they allow for the testing of a different ability of *DHR96* mutant flies – their ability to recover from drug treatment.

15 If, indeed, *DHR96* mutants express lower levels of detoxifying enzymes than wild type flies, a slower rate of recovery for mutant flies exposed to a drug should be seen. This test requires treating young adult flies with sub-lethal doses of a drug and then scoring the time it takes for those animals to regain normal behavior following transfer back to normal food. The choice of assay to measure behavior depends on the type of drug being tested, as described above. The  
20 advantage of a recovery test is that it may uncover more subtle effects on detoxification gene expression than could be detected by the acute tests described above. For example, whereas mutant and wild type flies might show a small difference in negative geotaxis when challenged with a particular drug, assaying for the ability of these two stocks to recover from drug treatment may significantly increase this difference.

25 336. The above assays are for testing the effect of xenobiotics on adult flies. Compounds can also be tested for their larvicidal effects by administering them in the food to staged populations of larvae (Grant et al. Bull. Envir. Contam. Tox. 69:35-40, 2002). *DHR96* and Canton S control flies are maintained on normal cornmeal/molasses agar supplemented with yeast. Egg lays are collected overnight from these stocks and used to inoculate fresh vials of  
30 food supplemented with a specific concentration of the drug to be tested. The drug are mixed with either Instant *Drosophila* Medium (Formula 4-24, Carolina Biological Supply) or added to a defined growth medium for *Drosophila* (Sang et al.). The Instant Medium is a flake formulation that is simply mixed with water before use. Drugs at different concentrations can be

easily added to each vial and mixed into an even suspension for oral delivery. The defined medium is in an agar base and thus the drug needs to be added as the food is being prepared. The advantage of the former is its ease of use. The advantage of the latter is its defined constitution of specific amino acids, vitamins, and other essential nutrients. The use of the Carolina Instant medium with drugs such as tebufenozide (Fig. 11C) has already been tested.

337. All studies described above are conducted with a *DHR96* mutant stock that has been outcrossed for 10 generations to the Canton S control stock. As a further test of specificity, toxin sensitivity rescue can be tested by using a wild type *DHR96* transgene in a *DHR96* mutant background. Two transgenes are used for this propose. First, the heat-inducible *hsp70-DHR96* fusion gene described above can be used. This construct has been established in transformed flies and used to overexpress wild type DHR96 protein (Fig. 10). This transgene has been crossed into a *DHR96* mutant background and expressed DHR96 protein with a 30 minute 37°C heat treatment. Western blots reveal that DHR96 protein can be easily detected at 24 hours after heat induction, at levels comparable to endogenous expression, indicating that the protein is relatively stable (Fig. 10). This *hsp70-DHR96* transgene can be crossed into the tenth outcross stock of the *DHR96*<sup>E25</sup> mutant and DHR96 expression induced by a single 30 minute 37°C heat treatment in larvae or adult flies tested with the drug. *DHR96* mutant and Canton S control animals are subjected to an identical heat treatment regime to control for any effects due to temperature. The appropriate drug and assay can then be used, as described above, to determine how the transgene affects the *DHR96* mutant phenotype. Thus, for example, while *DHR96* mutant flies might show sensitivity to a particular drug under conditions in which Canton S flies are relatively normal, this sensitivity can be rescued by heat-induced DHR96 expression, essentially recovering wild type function.

338. A second rescue construct can be used that does not depend on heat-induced expression. A 11.8 kb fragment, extending from 2.5 kb 5' of the wild type *DHR96* gene to 2.8 kb 3' of the gene, can be excised from a P1 genomic clone and inserted into the Carnegie 4 fly transformation vector (Rubin et al., Nucleic Acids Res 11:6341-51, 1983). This *DHR96* rescue fragment is introduced into the fly genome using standard methods for transformation, and crossed into the *DHR96*<sup>E25</sup> mutant background. Western blot analysis of this stock can reveal a recovery of wild type levels of DHR96 protein, indicating that the transgene is functioning as expected. This rescued stock, along with the *DHR96* mutant and Canton S control, can then be tested using an appropriate drug assay. Both the Canton S and rescued stock can show a similar

wild type response while the *DHR96* mutant shows a defective response, indicating that the phenotype seen in the mutant can be specifically ascribed to the *DHR96* locus.

339. Finally, it can be determined whether *DHR96* overexpression in a wild type genetic background has any effects on xenobiotic sensitivity. The *hsp70-DHR96* transgene is crossed into a Canton S background to ensure that no phenotypic differences between these stocks are due to genetic background. Heat-induced *hsp70-DHR96* transformants are then tested with a range of compounds, using assays as described above, comparing their sensitivity to heat-treated Canton S controls. This gain-of-function genetic test complements the loss-of-function genetics described above.

#### 4. Example 4: A role for *DHR96* in the regulation of specific detoxifying genes

340. Genes that are expressed in response to xenobiotic challenge can be identified, and it can be determined what role *DHR96* might play in mediating this regulation. The observation that *DHR96* mutants display a reproducibly increased sensitivity to phenobarbital (Fig. 11B) can be used. This compound has been used extensively in vertebrates for inducing xenobiotic responses and studying the transcriptional functions of the PXR and CAR xenobiotic receptors (Sueyoshi et al. Annu Rev Pharmacol Toxicol 41:123-43, 2001). Phenobarbital is also the most widely used inducer of xenobiotic gene transcription in insects. In *Drosophila*, it has been shown to have a significant effect on *Cyp6a2*, *Cyp6a8*, *Cyp6a9*, and *Cyp28* transcription, genes that are proposed to have xenobiotic activity. Northern blot hybridizations have been used to study the effects of phenobarbital on *Cyp6a2* and *Cyp6a8* transcription in wild type and *DHR96* mutant adult flies treated with 0.3%, 1%, and 3% phenobarbital. These results showed a dramatic induction of Cyp transcription in wild type animals, although no change in expression was seen in the *DHR96* mutant. As many potential detoxifying genes as possible can be considered. Canton S wild type and *DHR96*<sup>E25</sup> mutant adult flies, of identical genetic background and age, can be treated with either sucrose alone, or sucrose and 0.3% phenobarbital. This concentration is the lowest one at which *DHR96* mutants show a clear and reproducible sensitivity to the drug relative to wild type (Fig. 11B). It is also one that has been used in published studies of phenobarbital induced genes in *Drosophila* (Dunkov et al. DNA Cell Biol. 16:1345-56, 1997; Brun et al. Insect Biochem Mol Biol 26:697-703, 1996). Each treatment is done in triplicate. RNA is extracted from each set of animals, purified by TRIzol extraction (Gibco BRL) followed by RNeasy column chromatography (Qiagen), and ethanol precipitation. The RNA is then labeled and hybridized to Affymetrix GeneChip® *Drosophila* Genome 2.0

arrays designed to detect 18,500 *Drosophila* transcripts. Data is then analyzed using DChip 1.3 (<http://biosun1.harvard.edu/complab/dchip/>) and Significance Analysis of Microarrays (SAM). The data is scanned for changes in *Cyp6a2* and *Cyp6a8* mRNA levels, to confirm that phenobarbital treatment has had the expected effect in both wild type and *DHR96* mutant animals. *Cyp6a9* and *Cyp28* induction in wild type animals based on published data can also be seen (Danielson et al., Proc Natl Acad Sci 94:19797-802, 1997). Additional attention is paid to the genes that were identified by *DHR96* overexpression as potential regulatory targets.

341. There are two sets of data that emerge from this study. First, the data from untreated and treated Canton S controls identifies, for the first time, the genomic response to a xenobiotic compound in a wild type insect. This data can be analyzed to identify as many known detoxification genes as possible, focusing on the four main classes. Comparisons can be made with previous microarray studies that examined *Drosophila* genes involved in oxidative stress, to identify common stress response pathways (Landis et al. Proc Natl Acad Sci, 101:7663-8, 2004; Girardot BMC Genomics, 5:74, 2004). Gene ontology listings of array data can also be examined to identify new players in the xenobiotic response pathway (Misra et al. Genome Biol. 3:83, 2002). The second set of data to emerge from this microarray study allows for the determination of how *DHR96* might contribute to xenobiotic transcriptional responses in *Drosophila*. By comparing the set of genes regulated by phenobarbital in Canton S animals to those same genes in the *DHR96* mutant, it can be determined whether *DHR96* is required for this transcriptional response. Some genes can change their expression in wild type animals treated with phenobarbital will respond differently in *DHR96* mutants. The number and type of these gene changes provides insights into why *DHR96* mutants are more sensitive to phenobarbital than Canton S control animals. In addition, this experiment provides possible direct targets of *DHR96* transcriptional control, providing a foundation for the experiments described below.

342. Genes that change their regulation in Canton S animals treated with phenobarbital, and genes that are affected by the *DHR96* mutant, are validated by northern blot analysis. Collections of adult animals fed phenobarbital, as described above, can be used along with dose-response and time-course studies to understand the mechanisms of xenobiotic gene regulation. Validation can be conducted on selected genes, covering the different classes of detoxification pathways as well as new players that identified. Similar microarray studies using at least two other compounds, depending on which compounds show an effect in the viability and behavioral assays. It will be confirmed that wild type Canton S flies show a response to DDT using *Cyp12d1* and other P450 genes as probes for northern blot hybridization. One

experiment showed a low level of *Cyp6g1* induction by DDT in Canton S. Provided that a response can be detected, the survey can be conducted of DDT-regulated genes by performing microarray studies similar to those reported above for phenobarbital. Alternatively, it can be determined whether senita cactus alkaloids, compounds that have been shown to regulate the three *Cyp28* genes in *Drosophila mettleri*, also regulate these genes in *D. melanogaster* (Danielson et al. Proc Natl Acad Sci 94:10797-802, 1997). Other pesticides can also be surveyed for effects on a select group of *Cyp* gene targets to identify other compounds for use in comparative microarray profiling. The genomic response to these compounds can be determined and compared with the phenobarbital response, as well as determine how *DHR96* impacts these regulatory pathways. Determining the transcriptional response to more than one xenobiotic compound can provide an initial impression of how insects respond to different toxins in their environment. It is possible that a common core defense response can be activated in response to a range of drugs. Alternatively, the genetic response may be fine-tuned to combat specific xenobiotic compounds.

#### 5. Example 5: DHR96 activation by xenobiotic compounds

343. The human PXR xenobiotic nuclear receptor can directly bind xenobiotic compounds in its ligand binding pocket (Watkins et al., Science, 292:2329-2333, 2001), triggering induction of PXR targets, including the CYP3A detoxifying gene (Jones et al. Mol Endocrinol 14:27-39, 2000). This defines a positive feedback loop in which toxic compounds directly induce the expression of detoxifying genes through the PXR receptor. It can be determined whether DHR96 (the fly homolog of PXR, Fig. 1), acts in a similar manner. Several lines of evidence suggest that DHR96 might require a ligand for its activity. First, it is constitutively expressed throughout development, indicating that any temporal or spatial specificity for activation would have to be conferred post-transcriptionally. Second, ectopic overexpression of DHR96 has no effects on growth or development, unlike the majority of *Drosophila* orphan nuclear receptors that appear to act as constitutive transcriptional regulators (Thummel, Cell 83:871-7, 1995). Third, ectopic overexpression of DHR96 represses target genes, as shown by the microarray study (Fig. 12), similar to unliganded nuclear receptors such as the thyroid hormone receptor (Hu et al. Trends Endocrinol Metab 11:6-10, 2000). Finally, good evidence exists that the close relative of DHR96, the *C. elegans* DAF-12 receptor (Fig. 1A), is regulated by a steroid ligand (Matyash et al. PloS Biol. 2, e280, 2004, Gerisch et al. Development 129:1739-50, 2004).

344. DHR96 activation can be assayed for by using a method established to follow the activation status of a nuclear receptor ligand binding domain (LBD) in a developing animal. This method uses transformed *Drosophila* that carry the *hsp70* heat-inducible promoter upstream from the coding region for the yeast GAL4 DNA binding domain fused to the coding region for the DHR96 LBD (Fig. 13). These *hs-GAL4-DHR96* transformants are crossed with flies that carry a GAL4-dependent promoter driving a *lacZ* reporter gene that expresses nuclear  $\beta$ -galactosidase (*UAS-lacZ*). Expression of  $\beta$ -galactosidase can be detected by histochemical staining using X-gal as a substrate, generating a blue dye (Fig. 13, 14). A UAS-GFP reporter has also been used to detect GAL4-LBD activation in living animals, although this assay is somewhat less sensitive than that provided by  $\beta$ -galactosidase detection. The *hsp70* promoter was selected in order to provide precise temporal control, reducing potential lethality that might be caused by overexpression of the GAL4-LBD fusion protein (similar fusions to nuclear receptors have been shown to function as dominant negatives). In addition, the *hsp70* promoter should direct widespread expression of the GAL4-DHR96 protein upon heat induction, allowing for the assay for activation throughout the animal. Activation by this fusion protein, however, should only occur at times and in places where the appropriate hormonal ligand and/or co-factors are present. This method thus provides a visual readout of where and when an LBD can be activated in the context of an intact developing animal, providing a powerful tool for defining nuclear receptor signaling pathways. This system has been used to characterize the activation patterns of the *Drosophila* EcR and USP nuclear receptors, which act as a heterodimeric receptor for the steroid hormone ecdysone (Kozlova et al. 129:1739-1750, 2002). More recently, all 18 canonical *Drosophila* nuclear receptors have been used, defining their activation patterns during both embryogenesis and metamorphosis. These experiments have shown that GAL4-DHR96 is not normally active in wild type animals.

345. To test that, like its vertebrate counterparts, DHR96 is activated by xenobiotic compounds, thereby inducing the expression of detoxification target genes, activation of the GAL4-DHR96 fusion protein by xenobiotic compounds using three different means of compound delivery: (1) adding xenobiotic compounds to cultured third instar larval organs, (2) feeding larvae with xenobiotic compounds, and (3) feeding adult flies with xenobiotic compounds.

346. An advantage of the GAL4-LBD system is that it can be used in tissues dissected from transgenic larvae to test specific compounds for their ability to activate the fusion protein. Thus, for example, the steroid hormone 20-hydroxyecdysone is a potent activator of the GAL4-

USP fusion protein, and this response is dependent on its EcR partner, as expected (Kozlova et al. Development 129:1739-50, 2002). Similarly, tests of several compounds using the GAL4-LBD system in cultured larval organs revealed that the *Drosophila* NGFI-B ortholog, DHR38, can be activated by  $\alpha$ -ecdysone and 3-epi-20-hydroxyecdysone, but not 20-hydroxyecdysone. A similar assay can be used to test the ability of xenobiotic compounds to activate the GAL4-DHR96 fusion protein in cultured larval organs, using either *UAS-lacZ* or *UAS-GFP* as a readout. A few compounds have been tested in this manner in an initial effort to determine whether this approach will work as desired with the GAL4-DHR96 fusion. Of the compounds tested (DDT, phenobarbital, and tebufenozide), tebufenozide showed a reproducible and distinct pattern of activation. Control tissues dissected from heat-induced *UAS-lacZ* larvae treated with either vehicle alone or tebufenozide, or heat-induced *hs-GAL4-DHR96; UAS-lacZ* larvae treated with vehicle alone, gave a low background pattern of activation (control in Fig. 14). In contrast, larval organs dissected from *hs-GAL4-DHR96; UAS-lacZ* larvae and treated with tebufenozide gave a reproducible pattern of activation (GAL4-DHR96 in Fig. 14). Interestingly, this pattern is similar to that of endogenous DHR96 protein: in the fat body, midgut (but not restricted to the gastric caeca), and Malpighian tubules (but not salivary glands).

347. Organs isolated from other stages of development can be tested for their ability to direct GAL4-DHR96 activation by tebufenozide, to control for the possibility that a critical co-factor for DHR96 activation can be temporally restricted. The stage used for the experiment depicted in Fig. 14 is not ideal as mid- and late third instar larvae stop feeding in preparation for metamorphosis. Actively feeding stages during the second and early third instar can therefore be tested. Finally, it can be determined whether a natural form of compound delivery is more effective at revealing GAL4-DHR96 activation than using an *in vitro* organ culture system. Providing compounds to the animal in their growth medium allows for entry through the digestive system, epidermis, and/or tracheal system. Compounds added in this way can then have either a direct effect on the GAL4-DHR96 reporter or an indirect effect, with LBD activation occurring via a metabolic product of the compound being tested. Compounds are fed to control *UAS-lacZ* larvae and *hs-GAL4-DHR96; UAS-lacZ* larvae using either Instant *Drosophila* Medium (Formula 4-24, Carolina Biological Supply) or the defined growth medium. These animals are then be heat-treated, allowed to recover for 4-6 hours, and the patterns of *lacZ* expression are determined by Xgal assays (or fluorescence can be used to detect GFP for the *UAS-GFP* reporter gene). The methods described above can also be used to provide xenobiotics to adult *Drosophila*, feeding with a sucrose solution or using a contact assay. Taken together,

these assays should provide a list of compounds that can activate the GAL4-DHR96 LBD fusion protein in an intact animal, providing a basis for determining whether these compounds directly activate the DHR96 receptor as well as a means of understanding how xenobiotic compounds are sensed in insects.

5           348. While the GAL4-LBD system can be used to identify compounds that activate the LBD, it does not indicate the mechanism by which this activation is achieved. This effect could be obtained by direct binding of the compound to the LBD, as is the case for the EcR/USP heterodimer in *Drosophila*, or it could be due to the recruitment of protein co-factors or any post-transcriptional modification that could provide a transcriptional activation function.

10       Accordingly, compounds that are scored as positive by our GAL4-DHR96 assay act directly on the DHR96 LBD are tested.

#### **6. Example 6: Conserved regulatory sequences in detoxification target promoters.**

15           349. The studies described above provide insights into how xenobiotics are sensed by insects and how the animal reprograms its gene expression to detoxify these compounds. Biochemical techniques can be used to determine whether DHR96 functions as a monomer, homodimer, or heterodimer with USP, and determine its DNA binding specificity. Second, the

20       sequences bound by DHR96 can be tested *in vivo*, using chromatin immunoprecipitation (ChIP) and antibody stains of the larval salivary gland polytene chromosomes. Comparison of this data with the *in vitro* DNA binding results should provide an understanding of how DHR96 contacts target genes and identify potential regulatory targets in the genome for further characterization. Third, the regulatory sequences of coordinately expressed detoxification genes can be compared,

25       as determined by the microarray studies, to identify common sequence elements. It can be determined which of these sequence elements are bound by DHR96 and which might be bound by other regulatory factors. Taken together with the functional studies described herein, this work can provide a strong foundation for understanding how insects reprogram their patterns of gene expression to respond to toxic compounds in their environment.

30           350. DHR96 contains a novel P box sequence within its DNA binding domain: ESCKA (Fisk et al. Proc Natl Acad Sci, 92:10604-8, 1995). This P box is shared by only three other nuclear receptors in any organism – the three *C. elegans* homologs of DHR96: DAF-12, NHR-8, and NHR-48 – suggesting that DHR96 regulates a unique set of target genes in the insect genome. Consistent with this observation, it was found that DHR96 protein fails to bind

to most canonical nuclear receptor response elements, except for weak binding to a palindromic ecdysone response element (EcRE). A recent paper has determined the DNA sequences bound by DAF-12, providing initial insights into the binding specificity of this receptor subfamily (Shostak et al. Genes Dev 18:2529:44, 2004). They identified a direct repeat of two distinct hexanucleotide sequences (AGGACA and AGTGCA), separated by five nucleotides (DR5), as a functional DAF-12 binding site and response element. The authors proposed that DAF-12 would contact these sequences as a homodimer, although no experiments were done to address this issue. The DNA sequences bound by DHR96 can be determined. As a first step toward this goal, we will determine whether DHR96 acts as a monomer, a homodimer, or forms a heterodimer with USP, the fly ortholog of vertebrate retinoid X receptor (RXR). The vertebrate DHR96 homologs, PXR, CAR, and VDR, all act as heterodimers with RXR, suggesting that this interaction may have been conserved through evolution. Like vertebrate RXR, USP heterodimerizes with multiple nuclear receptor partners, including EcR and DHR38, indicating that it has relatively broad regulatory functions. GST-tagged USP protein are overexpressed in bacteria and purified by glutathione chromatography. All tags are added to the amino-terminal ends of the proteins, distant from the C-terminal dimerization sequences within the LBD. GST-USP is mixed with either FLAG-EcR or FLAG-DHR96, purified by glutathione chromatography, fractionated by gel electrophoresis, and FLAG-tagged proteins that are bound by GST-USP can be detected by Western blot analysis using anti-FLAG antibodies. Detection of the EcR/USP heterodimer acts as a positive control for this study. Results from this experiment can be confirmed by performing protein-protein interaction studies using either radiolabeled or unlabeled DHR96 and USP proteins synthesized *in vitro*, and our anti-DHR96 antibodies or AB11 mouse monoclonal antibodies directed against USP for immunoprecipitation. Again, detection of the EcR/USP heterodimer can be used as a positive control. These studies are directed at determining if DHR96 can heterodimerize with USP. To test if DHR96 can homodimerize, co-express GST-tagged DHR96 and FLAG-tagged DHR96 by *in vitro* translation. Protein is purified by using affinity beads for one of the two tags, and the presence of the other tag is assayed by gel electrophoresis followed by Western blot analysis, using antibodies directed against GST or anti-FLAG antibodies (both are commercially available).

351. To facilitate our identification of DHR96 regulatory targets, it can be determined which DNA sequences are preferentially bound by this transcription factor. DHR96 protein can be overexpressed and purified. This protein can be used either alone or in equimolar combination with purified USP, depending on whether it forms a USP heterodimer. USP is

purified from an overproducing strain of baculovirus, generously provided by M. Arbeitman and D.S. Hogness (Arbeitman et al. Cell 101:67-77, 2000). The selected and amplified binding site assay (SAAB) developed originally by Blackwell and Weintraub can be used. This method has been used widely to determine the optimal recognition sequences for DNA binding proteins. By using PCR to amplify each round of oligonucleotides that are selected for their ability to bind to DHR96, multiple random positions in the DNA sequence can be used, and thus better determined which sequences are optimally recognized by the protein. One choice of oligonucleotide sequences for this study can be informed by our earlier determination of how DHR96 contacts DNA, as a monomer, homodimer, or USP heterodimer. A pallindromic arrangement of random hexanucleotide sequences can also be tested, based on the identification of weak binding to the pallindromic EcRE, as well as a DR5 arrangement of hexanucleotide sequences based on the DAF-12 binding site. This analysis provides a set of ideal high affinity DHR96 binding sites, allowing for the determination of an optimal consensus recognition sequence. Although such ideal sites are rarely used *in vivo*, they nonetheless provide an invaluable guide for identifying *bone fide* binding sites within cis-acting regulatory sequences. For example, the determination of an optimal E74A ETS-domain DNA binding site by random oligonucleotide selection greatly facilitated the identification of downstream target genes (Urness et al. EMBO J 14:6239-46).

352. DHR96 binding sites used *in vivo* can also be used, and, by comparing them with the above biochemical data, define a set of potential direct regulatory targets in the genome. Two methods are used to determine where DHR96 protein is bound – antibody stains of the giant larval salivary gland polytene chromosomes and chromatin immunoprecipitation (ChIP). The giant larval salivary gland polytene chromosomes provide a unique and powerful tool for defining gene regulatory circuits in *Drosophila*. The fortuitous expression of DHR96 in the salivary glands of late third instar larvae provides an ideal opportunity to map its natural binding sites along the length of the giant polytene chromosomes. Since the cytological location of genes on the chromosomes has been well defined and correlated with the *Drosophila* genome sequence, DHR96 polytene binding sites can be matched to specific regions of DNA (Flybase Consortium, 2003 Nul Acid Res. 31:172-5). A similar genome-wide study of the *in vivo* binding sites of transcription factors has been conducted by using antibody stains of the polytene chromosomes, and these results have been used to predict direct regulatory targets which, in turn, have been confirmed at the molecular level. An advantage of this approach is that it is rapid, easy, and provides a complete survey of the genome. A clear shortcoming, however, is that this

method only allows a resolution of several hundred kilobases of genomic DNA. To overcome this problem, the search can be focused on binding sites on candidate genes that encode detoxification enzymes. Polytene binding data can be cross-referenced with the results of the microarray studies described above to identify likely DHR96 gene targets. These genes can be scanned for clusters of DHR96 binding sites, as determined by the biochemical studies described above. Finally, *in vivo* binding of DHR96 to specific sequences by ChIP is determined, as described below.

353. ChIP has been widely used to identify *in vivo* binding sites for DNA binding proteins, in many different organisms (Weinmann et al. Methods 26:37-47, 2002). Moreover, ChIP protocols are available for cultured cells, intact tissues, *Drosophila* embryos, or *Drosophila* adults, facilitating the use of this method (Cavalli et al., Damjanovski et al., Schwartz et al.). Two third instar larval tissues can be focused on, the fat body and salivary glands, both of which contain high levels of nuclear DHR96 protein. Crosslinking is performed using 0.3% formaldehyde, chromatin is fragmented by sonication, and aliquots are flash frozen in liquid nitrogen for subsequent chromatin immunoprecipitation. Efficient sonication of chromatin is tested by gel electrophoresis of purified DNA. DHR96 antibodies are used as a means of purifying chromatin fragments that are crosslinked to DHR96 protein. Antibodies effectively immunoprecipitate purified DHR96, and thus can work well for chromatin IP. If the antibodies fail to work as desired, affinity-purified and tested DHR96 antibodies from the antisera of two other rabbits can be used. Alternatively, if all antibodies fail, ectopically expressed tagged DHR96 can be used for chromatin IP. PCR can then be used to assay for the enrichment of DNA sequences that encompass potential DHR96 binding sites, as determined by biochemical studies described above as well as our polytene chromosome binding data. Attention can also be paid to promoters that are regulated by DHR96 as determined by microarray studies. Finally, potential DHR96 binding sites can be tested that are identified by bioinformatics, as described below.

354. In parallel with the above studies that are aimed at defining the DNA binding specificity of DHR96, conserved potential regulatory sequences can be determined within co-expressed target genes identified by the microarray studies. The microarray experiments described above generate two gene lists for each compound tested – one list showing which genes change their level of expression in response to a xenobiotic compound in wild type animals, and a second list showing which of those genes require *DHR96* for that regulatory response. These gene lists can be used to scan for clustered regulatory elements that are conserved between multiple co-regulated genes using several bioinformatic approaches. This

effort can identify novel DHR96 binding sites in the genome. In addition, other conserved regulatory elements can be determined that expands the understanding of detoxification gene expression beyond DHR96.

355. Bioinformatics is a rapidly evolving area with a number of labs developing and  
5 improving algorithms for mapping and predicting transcription factor binding sites. One  
program to identify nuclear receptor binding sites is “cis-analyst” (<http://rana.lbl.gov/cis-analyst/>). This is a web-based visualization tool that scans a given genomic region for the  
presence of a specific binding site consensus sequence, allowing the user to establish a cutoff  
point for eliminating weak binding sites. It searches for sequences of a specified length that  
10 contain a minimum number of predicted binding sites, allowing the detection of binding site  
clusters. This provides an ideal computational tool to enhance for functional sites rather than  
orphan binding sites that one might encounter on a random basis. The program generates a  
readily analyzed visual output that depicts binding sites on the DNA, along with genome  
annotation (Berman et al. Proc Natl Acad Sci, 99:757-62, 2002). Cis-analyst has been used to  
15 identify novel clustered binding sites for five well characterized *Drosophila* transcription factors,  
and these new regulatory targets have been validated by *in vivo* studies in transgenic animals  
MatInspector and Patch can also be used to look for binding sites of known transcription factors  
in *Drosophila* promoters of interest (<http://www.gene-regulation.com/pub/programs.html>), and  
Improbizer to scan for sequences that occur with an improbable frequency in a given segment of  
20 DNA (<http://www.cse.ucsc.edu/~kent/improbizer/improbizer.html>). These or similar programs  
can be used to analyze the promoter sequences of co-regulated genes identified by the microarray  
studies.

356. In order to determine whether the sequences identified above are likely to have  
functional significance, it can be determined if they have been conserved through *Drosophila*  
25 evolution. Evolutionary conservation has been widely used as a means of parsing regulatory  
sequences to identify true functional elements. This is particularly powerful in *Drosophila*,  
where the genome sequences of eight different species is becoming available. The first such  
sequence, that of *Drosophila pseudoobscura* (which diverged from *D. melanogaster* ~45 million  
years ago), was available earlier this year (<http://www.hgsc.bcm.tmc.edu/projects/Drosophila/>).  
30 This has now been supplemented with the ongoing genomic analysis of six other species,  
including *Drosophila virilis*, which diverged from *D. melanogaster* ~60 million years ago  
(<http://www.genome.gov/11008080>; <http://rana.lbl.gov/Drosophila/multipleflies.html>). The cis-  
regulatory sequences can be analyzed from selected detoxification target genes using as many of

these species as possible in order to determine whether DHR96 binding sites, or the binding sites of potential new transcriptional regulators, have been conserved through *Drosophila* evolution. Although confirmatory, this is an important step in determining whether the sequences we identify by informatics are likely to be functional *in vivo*.

5

#### 7. Example 7: The molecular mechanisms of detoxification gene expression.

357. The functional significance of these elements using both biochemical and genetic approaches can be determined. Nuclear extracts are prepared from larval fat bodies using published protocols (Lehmann et al. EMBO J 14:716-26, 1995; Antoniewski et al. Mol. Cell Biol 14:4465-74, 1994; von Kalm et al. EMBO J 13:3505-16, 1994). The choice of fat bodies derives from its functional equivalence to the mammalian liver as well as the abundant expression of DHR96 in this tissue. Sequences that encompass prospective DHR96 binding sites, or the binding sites of other potential regulators, are amplified by PCR and tested for their ability to be bound by factors in the fat body nuclear extracts. Protein binding to these fragments will be is monitored by electrophoretic mobility shift assays (EMSAs). The specificity of potential DHR96 interactions is determined by competition experiments using an oligonucleotide with an idealized DHR96 binding site, as well as by using DHR96 antibodies to supershift the complex. Antibodies directed against USP can be used to determine whether the binding complex also contains this potential heterodimer partner. Competition assays and antibody supershift experiments can be used to identify factors that bind to other conserved regulatory elements. The identity of some of these transcription factors, for example GAGA factor or C/EBP, should be predictable based on their DNA binding specificity (Lehmann et al., Park et al. DNA Cell Biol. 15:693-701, 2004). Other potential regulators can be found based on the sequences of oligonucleotides that efficiently compete for binding in nuclear extracts, and confirm this deduction by using appropriate antibodies for supershift studies. This approach has been used to identify ecdysone-regulated transcription factors that control glue gene transcription in *Drosophila* salivary glands as well as characterize ecdysone-inducible *Fbp-1* transcription in fat bodies.

358. The above studies confirms the presence of functional DHR96 binding sites in target promoters as well as allows for the identification of other potential trans-acting regulators of detoxification gene expression. The corresponding sequences in the target promoters are disrupted by site-directed mutagenesis using PCR. The resultant mutated fragments are tested by DNA sequencing to ensure that only the desired base changes have occurred. These fragments

are then be tested by EMSA to confirm that the mutations have disrupted binding to the corresponding transcription factor. The mutated fragments are then be used in combination with wild type sequences to reassemble target promoters for functional studies in transgenic animals.

359. Studies can also be conducted in transgenic animals as a means of determining  
5 the functional significance of specific transcription factor binding sites. 2-3 target promoters can be defined in the preceding specific aim, but can include other promoters to test specific hypotheses regarding possible transcription factor interactions that arise. Each of the target promoters can be fused to a *lacZ* reporter gene in the P element transformation vector pCaSpeR-AUG- $\beta$ gal (Thummel et al. Dros. Info. Services 71:150, 1992). These are introduced into the fly  
10 genome using conventional methods and multiple independent insertions are isolated to control against the effects of flanking sequences on reporter gene expression. Each promoter-*lacZ* fusion transgene is crossed into wild type and *DHR96* mutant genetic backgrounds to establish permanent stocks. These animals are exposed to either regular food or food supplemented with a xenobiotic, after which dissected tissues are tested for  $\beta$ -galactosidase expression using X-gal  
15 staining. Responses to phenobarbital can be tested based on earlier studies which showed that several hundred base pairs of the *Cyp6a2* or *Cyp6a8* promoter is sufficient to mediate phenobarbital-inducible transcription of a reporter gene in transgenic wild type *Drosophila*. Little or no  $\beta$ -galactosidase expression can be seen in tissues dissected from untreated wild type animals, and high levels of  $\beta$ -galactosidase expression in tissues from wild type animals exposed  
20 to phenobarbital. X-gal assays are performed on tissues dissected from *DHR96* mutant animals.

360. The wild type promoter sequences in the transgene vectors can be replaced with the mutated fragments described above, and introduce these P elements into the genome of both wild type and *DHR96* mutant animals. As before, multiple independent transgenic lines can be established to control against the effects of flanking sequences on reporter gene expression. The  
25 regulation conferred by the mutant promoter fragment will be tested in transgenic animals after exposure to phenobarbital or other xenobiotics, depending on our earlier studies. If a reduction or absence of *lacZ* transcription is seen, then the regulatory interaction disrupted by the promoter mutation is of functional significance. Alternatively, no effect on *lacZ* transcription indicates that the binding site is not essential for proper promoter regulation. In this case, additional  
30 transgenic lines will be established that carry multiple binding site mutations for that transcription factor, to determine whether they act in a redundant manner. Similarly, the contributions of individual binding sites are tested in other transgenic lines.

361. The effects of mutations in DHR96 binding sites should confirm the studies of the wild type transgene in *DHR96* mutant animals. That is, if the wild type promoter is unable to respond to a xenobiotic in a *DHR96* mutant background, then that same promoter carrying mutated DHR96 binding sites should show defective xenobiotic responses in wild type animals.

5 A similar approach can be used to test the functional significance of other transcription factor binding sites, crossing wild type promoter-*lacZ* fusion transgenes into stocks that carry mutations in putative trans-acting regulators, combined with studies of promoter transgenes that carry mutations in the corresponding binding sites. Such a demonstration of both cis and trans effects can be taken as a good indication that the corresponding transcription factor is involved in the  
10 observed regulatory interaction. Methods are available that allow us to create clones of mutant tissue, so that the effects of otherwise lethal transcription factor mutations can be studied. Taken together, these studies of wild type and mutated promoter-*lacZ* transgenes should allow for the decoding of the mechanisms of detoxification gene expression. It can be determined which binding sites are critical for the activity of a specific detoxification gene promoter, and which  
15 binding sites mediate xenobiotic-inducible transcription. In addition, it can be determined which transcription factors act through these sequences as well as how these transcription factors might interact to control the xenobiotic response.

362. Disclosed are methods for screening for the presence of xenobiotic receptor ligands using the constructs and methods disclosed herein, such as those for the GAL4-DHR96  
20 fusions.

### G. References

363. Boyd, L., O'Toole, E. and Thummel, C.S. (1991). Patterns of E74A RNA and protein expression at the onset of metamorphosis in *Drosophila*. *Development* 112, 981-995.

364. Kozlova, T. and Thummel, C.S. (2002). Spatial patterns of ecdysteroid receptor  
25 activation during the onset of *Drosophila* metamorphosis. *Development* 129, 1739-1750.

365. Kozlova, T., and C.S. Thummel (2003) Methods to characterize *Drosophila* nuclear receptor activation and function in vivo. In: "Methods in Enzymology. Nuclear Receptors, Vol. 364 (Russell, D.W., and Mangelsdorf, D.J., eds.), Academic Press, New York, pp. 475-490.).

30 366. Rong, Y.K., Titen, S.W., Xie, H.B., Golic, M.M., Bastiani, M., Bandyopadhyay, P., Olivera, B.M., Brodsky, M., Rubin, G.M., and Golic, K.G. (2002). Targeted mutagenesis by homologous recombination in *D. melanogaster*. *Genes and Development* 16, 1568-1581.

367. Daborn, P. J., Yen, J. L., Bogwitz, M. R., Le Goff, G., Feil, E., Jeffers, S., Tijet, N., Perry, T., Heckel, D., Batterham, P., *et al.* (2002). A single p450 allele associated with insecticide resistance in *Drosophila*. *Science* 297, 2253-2256.

368. Danielson, P. B., Foster, J. L., McMahon, M. M., Smith, M. K., and Fogleman, J. C. (1998). Induction by alkaloids and phenobarbital of Family 4 Cytochrome P450s in *Drosophila*: evidence for involvement in host plant utilization. *Mol Gen Genet* 259, 54-59.

369. Danielson, P. B., MacIntyre, R. J., and Fogleman, J. C. (1997). Molecular cloning of a family of xenobiotic-inducible drosophilid cytochrome p450s: evidence for involvement in host-plant allelochemical resistance. *Proc Natl Acad Sci U S A* 94, 10797-

10802.

370. Fogleman, J. C. (2000). Response of *Drosophila melanogaster* to selection for P450-mediated resistance to isoquinoline alkaloids. *Chem Biol Interact* 125, 93-105.

371. Francis, G. A., Fayard, E., Picard, F., and Auwerx, J. (2003). Nuclear receptors and the control of metabolism. *Annu Rev Physiol* 65, 261-311.

372. Kliewer, S. A. (2003). The nuclear pregnane X receptor regulates xenobiotic detoxification. *J Nutr* 133, 2444S-2447S.

373. Li, C., and Wong, W. H. (2001). Model-based analysis of oligonucleotide arrays: model validation, design issues and standard error application. *Genome Biol* 2, RESEARCH0032.

374. Lindblom, T. H., Pierce, G. J., and Sluder, A. E. (2001). A *C. elegans* orphan nuclear receptor contributes to xenobiotic resistance. *Curr Biol* 11, 864-868.

375. Maglich, J. M., Stoltz, C. M., Goodwin, B., Hawkins-Brown, D., Moore, J. T., and Kliewer, S. A. (2002). Nuclear pregnane x receptor and constitutive androstane receptor regulate overlapping but distinct sets of genes involved in xenobiotic detoxification. *Mol*

*Pharmacol* 62, 638-646.

376. Makishima, M., Lu, T. T., Xie, W., Whitfield, G. K., Domoto, H., Evans, R. M., Haussler, M. R., and Mangelsdorf, D. J. (2002). Vitamin D receptor as an intestinal bile acid sensor. *Science* 296, 1313-1316.

377. Mangelsdorf, D. J., Thummel, C., Beato, M., Herrlich, P., Schutz, G., Umesono, K., Blumberg, B., Kastner, P., Mark, M., Chambon, P., and *et al.* (1995). The nuclear receptor superfamily: the second decade. *Cell* 83, 835-839.

378. Maurel, P. (1996). Cytochromes P450: Metabolic and Toxicological Aspects. In Cytochromes P450: Metabolic and Toxicological Aspects, C. Ioannnides, ed. (Boca Raton, CRC Press), pp. 241-270.

379. Ranson, H., Claudianos, C., Ortelli, F., Abgrall, C., Hemingway, J., Sharakhova, M. V., Unger, M. F., Collins, F. H., and Feyereisen, R. (2002). Evolution of supergene families associated with insecticide resistance. *Science* 298, 179-181.

380. Rong, Y. S., and Golic, K. G. (2000). Gene targeting by homologous recombination in *Drosophila*. *Science* 288, 2013-2018.

381. Rong, Y. S., Titen, S. W., Xie, H. B., Golic, M. M., Bastiani, M., Bandyopadhyay, P., Olivera, B. M., Brodsky, M., Rubin, G. M., and Golic, K. G. (2002). Targeted mutagenesis by homologous recombination in *D. melanogaster*. *Genes Dev* 16, 1568-1581.

382. Robinson-Rechavi M, Escriva Garcia H, Laudet V., The nuclear receptor superfamily, *J Cell Sci.* 2003 Feb 15;116(Pt 4):585-6

383. Tontonoz, P., and Mangelsdorf, D. J. (2003). Liver X receptor signaling pathways in cardiovascular disease. *Mol Endocrinol* 17, 985-993.

384. Wilson, T. G. (2001). Resistance of *Drosophila* to toxins. *Annu Rev Entomol* 46, 545-571.

## H. Sequences

### 1. SEQ ID NO: 1 Accession No. NM\_130611 *Drosophila melanogaster* CG16902-PA

MTLSRGPYSELDKMSLFQDLKLKRRKIDSRCSSDGESIADTSTS  
 SPDLLAPMSPKLCDSGSAGASLGASLPLPLALPLPMALPLPMSLPLPLTAASSAVTVS  
 25 LAAVVAAVAETGGAGAGGAGTAVTASGAGPCVSTSSTTAAAATSSSTSSLSSSSSSSSSS  
 TSSSTSSASPTAGASSTATCPASSSSSSSGNGSGGKSGSIKQEHTIHSSSSAISAAAA  
 STVMSPPPAEATRSSPATPEGGGPAGDGSGATGGGNTSGGSTAGVAINEHQNNNGNGSG  
 GSSRASPDSELEKPTTTTTTGRPTLTPTNGVLSSASAGTGISTGSSAKLSEAGMSVIR  
 SVKEERLLNVSSKMLVFHQREQETKAVAAAAAAAAGHVTVLVTPSRIKSEPPPPAS  
 30 PSSTSSTQRRERERDRERDRERERERDRDREREREQSISSSQQHLSRVASAPPTQLS  
 HGSLGPNIVQTHHLHQQLTQPLTLRKSSPTEHLLSQSMQHLTQQQAIHLHLLGQQQ  
 QQQQASHPQQQQQQQHSPHSLVRVKKEPNVGRHLSPHHQQQSPLLQHHQQQQQQQQQ  
 QQQHLHQQQQQQQQHQQQPQALALMHPASLALRNSNRDAAILFRVKSEVHQQVAAGLP  
 HLMQSAGGAAAAAAAVAAQRMVCFSNARINGVKPEVIGGPLGNLRPVGVGGGNGSGS  
 35 VQCPSHPSSSSSSSQLSPQTPSQTTPRGTPTVIMGESCGVRTMVWGYEPPPPSAGQS  
 HGQHPQQQQQQSPHHQPQQQQQQQQQQQQSQQQQQQQQQQQSLGQQQHCLSSPSAGSLTPSS  
 SSGGGSVSGGVGGPLTPSSVAPQNNEEAQLLSLGQTRIQDMRSRPHPFRTPHALN  
 MERLWAGDYSQLPPGQLQALNLSAQQQQWGSSNSTGLGGVGGGMGGRNLEAPHEPTDE  
 DEQPLVCMICEDKATGLHYGIITCEGCKGFFKRTVQNRRVYTCVADGTCEITKAQRNR  
 40 CQYCRFKKCIEQGMVLQAVREDRMPGGRNSGAVYNLYKVKYKKHKKTNQKQQQQAQQ  
 QQQQAAAQQQHQQQQQHQQHQQHQQQLHSPLHHHHHQGHQSHHAQQQHHPQLSPHHL  
 LSPQQQQLAAAVAAAAAQHQQQQQQQQQQQQQQAQKLMGGVVDMMKPMFLGPALKPELLQAP

PMHSPAQQQQQQQQQQQQQASPHLSLSSPHQQQQQQQGHQNHQQQQGGGGGGAGGG  
 AQLPPHLVNGTILKTALTNPSEIVHLRHRLDSAVSSSKDRQISYEHALGMIQTLIDCD  
 AMEDIATLPHFSEFLEDKSEISEKLCNIGDSIVHKLVS WTKKLPFYLEIPVEIHTKLL  
 TDKWHEILILTTAAYQALHGKRRGEGGSRHGSPASTPLSTPTGTPLSTPIPSPAQPL  
 5 HKDDPEFVSEVNSHLSTLQTCLTTLMGQPIAMEQLKLDVGHMVDKMTQITIMFRRIKL  
 KMEEYVCLKVYILLNKGTWFDLQNPFIQCSCYLLVRFVNPAEVELESIQERYVQVLRS

YLQNSSPQNPQARLSELLSHIPEIQAAASLLLESKMFYVPFVLNSASIR  
 ORIGIN

10 **2. SEQ ID NO: 2 Accession No. NM\_130611 Drosophila melanogaster**  
**CG16902-PA**

1 atgacactga gccgtggccc gtacagcgag ctcgataaaa tgagcctttt tcaagacctc  
 61 aaactcaaac ggcgcaaaat cgattcgca tgcagcagtg acggcgagtc catagcggac  
 15 121 acgtccacct cgtcgccgga cctgctggcg cccatgtcgc cgaagctctg cgacagcggc  
 181 tcggcggggg cgtcgctggg ggcategtg cccctgccgc tggcctgcc cctgccaatg  
 241 gccctgccac tgcccatgtc gctgccctg cccctcacgg cggcatctc ggcggtcacc  
 301 gtttcgtgg cagcggctgt ggccgcggtg gccgagacgg gtggcgcggg cgcgggagga  
 361 gctgggacag cagtaacagc gtcgggagca ggaccatgcg tctccacgtc gtctacgacg  
 20 421 gcagcggcag ccacatctc gacctctcg ctctgtctt cctctcttc gtcactctcc  
 481 acgtctcca gcacttctc cgctcgccg acagctggag cctctccac ggccacctgc  
 541 cccgccagca gcagcagcag cagtggaaac ggaagtgggg gcaaaagtgg tagcatcaag  
 601 caggagcaca cggagataca ctctcgagc agtgcgatt cggcgggcgc cgcctcaacg  
 661 gtgatgtcac cggcgccgc tgaggcgacg agatccagtc cagccacgcc cgaggagggc  
 25 721 ggaccagctg gcgacggaag tggagcaacg ggaggcgga acacgagcgg cggatcaacg  
 781 gctggagtgg ccattaatga acacaaaac aatggcaatg gcagcggcgg gagcagtcga  
 841 gcctctccg attcgctgga agagaagccc tctaccaca cgaccacagg tctccaacg  
 901 ctacgcccga cgaatgggt gctgtctcc gcctcgccg gcacggggat ttccacagga  
 961 agcagcgcca agctgagcga ggctggatg agtgatgac ggtccgtgaa ggaggagcgc  
 30 1021 ttgtcaacg tatccagcaa gatgctggtg ttcatcagc agcgggagca agagacaaaa  
 1081 gcagtggcgg ctgcagcagc agcagcagc gcgggcatg tgacggtct agtgacgcca  
 1141 tcgcatca aatcgagcc accgccgcg gcttaccct cctctacac cagcacaca  
 1201 agggaaaggg aacgggaacg cgtcgagag agggatcgc aaagggaacg cgagcgggac  
 1261 cgggaccggg aacgggaacg ggaacagtc atcagctct cgcagcagca cctaagtcgg  
 35 1321 gtctccgcca gtccaccac tcagctgtc cagggcagcc tgggaccaa cattgtgag  
 1381 acgcaccatc ttaccagca actcacacag ccgtgacgc tgcgaagag cagccgccc  
 1441 acagagcacc tgctcagtc gtccatgcaa catctcac agcagcaggc gatccacctg  
 1501 catcacctac ttggccagca gcagcagcag cagcagcgt cgcacccca gcagcaacag  
 1561 cagcagcaac actcgccca ctccctggtg cgggtgaaa aggaaccgaa tgttggtcag  
 40 1621 cggcacttat cgccgatca ccaacaacag tcgccactcc tgcagacca ccaacagcag  
 1681 cagcagcagc aacaacaaca gcaacagcat ctgcacagc aacagcaaca gcagcagcat  
 1741 caccagcagc agccccaggc actggccctg atgcacccg ctccctggc gctaaggaa  
 1801 agcaatcggg atgcggccat tctgttcgg gtgaagagc aagtgcacca gcaggtggcc  
 1861 gccgggtgc cgcactgat gcagtcgct ggtggggcag cggccgccgc cgcagcagct  
 45 1921 gtggccgctc agcgaatggt atgcttcagc aatgccagga tcaatggcgt taagccggag  
 1981 gtgattggag gaccgtggg caacctgcgg cccgtggcg tgggtggcg aaacggaagt  
 2041 ggctccgtgc agtccccct gccgcatca tctctctgt cgtcactc gcagctgtcg  
 2101 ccgcagcgc cctcccagac gccgccccga ggcacgcca ccgtcataat gggcgagagc  
 2161 tgcggggtgc gcaccatggt ctggggctac gacctccgc caccctcggc gggccagtcc  
 50 2221 cagggccagc acccgcaaca gcaacagcag tcgccccacc accagccgca acaacaacag  
 2281 cagcagcaac aacagcagtc gcagcagca cagcaacagc agcagcaaca gtcgtgggc  
 2341 cagcagcagc actgcctctc ctgccgtcg gcgggatcg tgacgccctc ctctctgtc  
 2401 ggcggtggtt cggtatctg cggcgagtg ggcggaccac tcacaccctc ctctgtgctg  
 2461 ccgcagaata acgaggaggc cggccaactc ctgtctccc tgggacagac acgcatccag  
 55 2521 gacatgagat cagggccaca cccctccgc acaccgacg ccttaatat ggagcggctg  
 2581 tgggcgggag actactcgca attgccgccc ggccagctgc aggtctgaa tctcagtgc

2641 caacagcagc agtggggcag cagcaactcc acgggtcttg gtggcgtagg cggcggcatg  
 2701 ggcgagacga acctggaggc gccgcacgag ccgaccgacg aggacgaaca gccgctcgtt  
 2761 tgcattgatct gcgaggacaa ggccaccggc ctgcactacg gcatcatcac ctgcgagggg  
 2821 tgcaagggct tctcaagcg gacgggtcag aaccgacgag tctacacctg cgtggcggac  
 5 2881 ggcacctgcg agataaccaa agcacagcgc aaccgttgc agtattgtcg attaagaag  
 2941 tgcattgagc agggcatggt gctgcaagcc gtgcgagag atcgcatgcc gggcggtcgc  
 3001 aacagtggcg ccgtctaca tttgtacaag gtgaagtaca agaagcaca gaagaccaat  
 3061 cagaagcagc agcagcaggc cggccagcag cagcagcagc aggcggcggc gcagcagcag  
 3121 caccagcaac agcagcagca tcaacagcac cagcaacatc agcaacagca gttgcactcg  
 10 3181 ccgtccacc atcaccacca ccagggccac cagtcgcacc acgcgcagca gcagcaccac  
 3241 ccacagctgt cggcgacca cctgctgtcg ccgcagcagc agcaacttgc cggcgcggtg  
 3301 gcagcagctg cgcagcacca acagcaacag caacaacagc agcaacagca gcagcaggcc  
 3361 aagctgatgg gcggcgtggt ggacatgaag cccatgttcc tcggccccgc ttgaagccg  
 3421 gattgtctgc aagcaccgcc catgcacagt cggcccgagc aacaacaaca gcagcagcag  
 15 3481 cagcagcagc aacagcaggc ctgcgcgat ctctcgctta gtcaccgca ccagcagcag  
 3541 cagcagcagc agggacagca ccaaaaccac caccagcaac aaggtggggg tggcggagga  
 3601 gctggtggag gagtcaact gccgcgcac ctggtgaacg gaacgatact gaagacggcc  
 3661 ctaaccaatc ccagcgagat tgtacatctg cgccaccgcc tcgactcggc ggtcagttcg  
 3721 tccaaggacc gacagatctc gtacgagcac gccttaggca tgatccagac actgatcgac  
 20 3781 tgcagcga tggaggacat agccacactg ccgcacttca gcgagttcct tgaggacaag  
 3841 tcggagatta gcgagaaact gtgcaacatc ggcgattcca tagtcacaa gctggtgtcg  
 3901 tggacaaaa agttgccctt ctacctggag atcccgggtg agatacatc caaactactg  
 3961 acggacaagt ggcacgagat ccttatcctg accacggccg cctaccaggc gttgcatggc  
 4021 aagcggcgtg gcgagggagg aggcagcagg catggttcgc cggcgtaac gccactgagc  
 25 4081 acgcccactg gtacgccgtt gagcacaccg ataccctcgc ccgcccagcc actgcacaag  
 4141 gacgacccgg agtttgcag cgaggtgaac tcgcacctga gcacactgca aacctgctt  
 4201 accacgctaa tgggccagcc gatagcgatg gagcagctga agctggacgt cgggcacatg  
 4261 gtggacaaga tgaccagat caccatcatg ttccggcgaa tcaagctcaa gatggaggag  
 4321 tacgtctgcc tgaaggtta catactgcta acaaaggta cgtggttga ttgcaaaac  
 30 4381 ccattcatac agtgcctatg ttacctctc gttcgtttg taaatccagc agaagtggaa  
 4441 ctggagagca tccaggagcg gtacgtccag gtgctgcgt cctacctgca aaactcctc  
 4501 ccgcagaatc cgcaggcgag gctcagtga ctgctctccc acataccaga gatccaggct  
 4561 gcggctagcc tgctgctcga gagcaagatg ttctatgtgc cctcgtgct caactcggcg  
 4621 agcataaggt ag

35

### 3. SEQ ID NO: 3 Accession No. NM\_168775 *Drosophila melanogaster* ftz transcription factor 1 CG4059-PA

MLEMDQQQATVQFISSLNISPFMSMQLEQQQQPSSPALAAGGNS  
 40 SNNAASGSNNNSASGNNTSSSSNNNNNNNDNDNAHVLTKEHEYNAYTLQLAGGGGSG  
 SGNQQHHSNHSNHGHHQQQQQQQQQQHQQQQQEHYQQQQQQNIANNANQFNSSSY  
 SYTYNFDQSQYIFPTGYQD TTSSHSQQSGGGGGGGGNLLNGSSGGSSAGGGYMLLPQA  
 ASSSGNNGNPAGHMSSGSVGNNGSGGAGNGGAGGNSGPGNPMGGTSATPGHGGEVIDF  
 KHLFEELCPVCGDKVSGYHYGLLTCECKGFFKRTVQNKKVYTCVAERSCHIDKTQRK  
 45 RCPYCRFQKCLEVGMKLEAVRADRMGRGNKFGPMYKRDRARKLQVMRQRQLALQALR  
 NSMGPDIKPTPISPGYQQA YPNMNIKQEIQIPQVSSLTQSPDSSPSPIAIALGQVNAS  
 TGGVIATPMNAGTGGSGGGGLNGPSSVGNNGNSSNGSSNGNNNSSTGNGTSGGGGGGNA  
 GGGGGGTNSNDGLHRNGGNGNSSCHEAGIGSLQNTADSKLCFDSGTHPSSTADALIEP  
 LRVSPMIREFVQSIDDREWQTQLFALLQKQTYNQVEVDLFELMCKVLDQNLFSQVDWA  
 50 RNTVFFKDLKVDDQMKLLQHSWSDMLVLDHLHHRHNLPLDETQLNNGQVFNLM SLGL  
 LGVPQLGDYFNLQNKLDLKFDMGDYVCMKFLILLNPSVRGIVNRKTVSEGH DNVQA  
 ALLDYTLTCYPSVNDKFRGLVNILPEIHAMAVRGEDHLYTKHCAGSAPTQTLLMEMLH

AKRKG

55

**4. SEQ ID NO: 4 Accession No. NM\_168775 Drosophila melanogaster ftz  
transcription factor 1 CG4059-PA**

1 ctacgcaaaa taaaacgtac atgaaatgtt attagaaatg gatcagcaac aggcgaccgt  
 5 61 acagtttata tcgtcgtga atatatgcc gttcagcatg cagctggagc agcagcagca  
 121 gccctccagt cccgctctgg ccgccggtgg caacagcagc aacaacgcgg ccagcggttag  
 181 caacaacaac agcgccagcg gcaacaacac cagcagcagc agcaacaaca acaacaaca  
 241 taacaacgac aatgatgcac acgttctaac gaaattcgag cacgaataca atgcctacac  
 301 gttgcagttg gccggaggcg gtgggagtgg cagcggcaat cagcagcacc acagcaacca  
 10 361 cagcaaccac ggcaaccacc accagcagca gcagcaaca cagcaacagc agcagcaaca  
 421 tcagcagcag cagcaagaac actaccagca gcaacagcaa cagaatatcg ccaacaatgc  
 481 caatcaatc aactcctcgt cctactcgt tataataat ttcgattcac agtatatatt  
 541 cccgacaggc taccaggaca ccacctcctc acactcgcaa cagagcggag gaggcggtgg  
 601 cggcggcggt ggcaacctgc taaacggcag ctccggcggc agctccgccg gcggtggcta  
 15 661 catgtctctc cccaggcgcc ccagctccag tggcaataat ggcaatccga atgccggcca  
 721 catgtctctc ggttccgtgg gcaatggcag cggaggcgct ggcaatggcg gagcggcgcg  
 781 caactccggt cccggcaatc ccatggcgcg tacgagcgcc acgccgggac acggcgcgga  
 841 ggtgatcgac ttaagcacc tgttcgagga gctttgccc gtgtgtggcg acaaggtag  
 901 cggctaccac tacggcctgc tcacctgca gtctgcaag ggattcttca agcgcaccgt  
 20 961 gcagaacaag aaggtctaca cctgcgtggc ggagcggctg tgccacatcg acaagacgca  
 1021 gcgcaagcgg tgcctact ggcgattcca gaagtgcctc gaggtgggca tgaagctaga  
 1081 ggctgttcca gcgtagaga tgcgtggtgg acgcaacaaa ttcggacca tgtacaaacg  
 1141 ggatcgcgcg cggaagtgc aagtatgcg gcagcggcag ttggcgctgc aagcgtgcg  
 1201 caactcgatg ggtccggaca tcaagccaac gccgatctcg ccgggctacc agcaagcata  
 25 1261 tccaaatatg aacattaagc aggaaattca aatactcag gtatctcac tcaccaatc  
 1321 tccggactcg tcgccagcc ccatagcaat tgcgttggga caggtgaacg cgagcacggg  
 1381 cgggtgttata gccacgcca tgaacgccgg cactggcggc agtggggcg gtggtctgaa  
 1441 cggaccaagt tccgtgggca acggcaatag cagcaacggc agcagcaacg gcaacaaca  
 1501 cagcagcacg ggcaacggaa cgtccggagg aggaggtggc aataatgcgg gcggcggagg  
 30 1561 aggaggaacc aattccaacg atggcctgca tcgcaacggc ggcaatggca acagcagttg  
 1621 ccacgaggct ggaataggat ctctcgagaa cacggccgac tcgaaattgt gcttcgattc  
 1681 tggcacacat ccatcgagca cagccgacgc gctaactgag ccattaagag tctaccgat  
 1741 gattcgtgaa tttgtgcaat ctattgacga tcgggaatgg cagacgcaac tgtttgcct  
 1801 gctgcagaag caaacctaca accaggtgga agtggatctc ttcgagctga tgtgcaaagt  
 35 1861 gctcgaccag aattgttct cgcaagtaga ctgggcacgg aacaccgtct tcttcaagga  
 1921 tctgaaggtc gacgacaaa tgaagctgt gcagcattcc tggcggaca tgcgttct  
 1981 ggatcacctg catcatcgaa tccataacgg cctgcccagc gagacgcaac tgaacaatgg  
 2041 tcaggtgttc aatctgatga gtctgggttt gttgggagtg ccacagctgg gcgattactt  
 2101 caacgagctg cagaacaagc tgcaggacct gaaattcgat atgggcgact atgtctgcat  
 40 2161 gaaattccta atcctgttga atccaagtgt acggggatt gtcaaccgga agaccgtctc  
 2221 cgagggacat gataatgtgc aagccgcttt gctggactac accctcacct gctatccgtc  
 2281 agtgaatgac aaattcagag ggctagttaa catcttaccg gaaatccatg ccatggccgt  
 2341 tcgcgcgag gatcacctgt acaccaagca ctgtgccggc agtgcgcca cccaaacgt  
 2401 gctcatggag atgctgcacg ccaagcgcaa gggatagagg ccgggagaac gtgacacgga  
 45 2461 atacttaatc atttatgaaa tgaataaac aaggcgggaa ggccctcggg gcaaccgggt  
 2521 catggaaggc gaacgaagga tacagcagaa ttccgtatta tgaatatggg aatgcatcat  
 2581 cactactacc accaactatc acacctatac acacacatgc acacattgt tgattcaatg  
 2641 ttaattatta ttacgttac ggtaggtct agtttacgt taactaatta attaatgt  
 2701 cttaaattaa ttcgtgtttt attttagtc cctgataaag caattttaa acactgaac  
 50 2761 ctaaacgaga atatgtagta gatgtatgga tttaattta aatacggcaa ggagaaacac  
 2821 acttttttag gcattacaaa aaaaagaag catgagaaat ttattttta taccctata  
 2881 tgaatcagat acttatggat acaaatctat atatatttt atgtaaattg gcgtacttt  
 2941 agcgtcctac atattttta attagaattt gggtatacta tagtttgaa attagtatcg  
 3001 tcccacttg aagatcgatt cttgtattt ttgcgcaa gtgtctgca tagtattgc  
 55 3061 gtctaataa atggcaacaa aaaaatatt ggaaatcca tacaagaaa atgaaaacaa  
 3121 agcaaattta ggtgtcatg gtagaatgt atgtgtatat tataattgta atttcatcta

3181 agtgaagaa aacaatgcaa acaactacct acaacaagat aatgaagagc aagaaattat  
 3241 ataaattaat aaaggctgtg ttaaaaact

## 5. SEQ ID NO: 5 Accession No. NM\_176123 *Drosophila melanogaster*

### 5 Hormone receptor-like in 46 CG33183-PA

MYTQRMFDMWSSVTSKLEAHANNLGQSNVQSPAGQNNSSGSIKA  
 QIEIIPCKVCGDKSSGVHYGVITCEGCKGFFRRSQSSVVNYQCPRNKQCVVDRVNRNR  
 CQYCRLQKCLKLGMSRDAVKFGRMSKKQREKVEDEVRFHRAQMRAQSDAAPDSSVYDT  
 10 QTPSSSDQLHHNNYNSYSGGYSNNEVG YGSPYGYASVTPQQTMQYDISADYVDSTTY  
 EPRSTIIDPEFISHADGDINDVLIKTLAEAHANTNTKLEAVHDMFRKQPDVSRILYYK  
 NLGQEELWLDCAEKL TQMIQNIIEFAKLIPGFMRLSQDDQILLKTGSFELAIVRMSR  
 LLDLSQNAVLYGDVMLPQEAFYTS DSEEMRLVSRIFQTAKSIAELKLTETELALYQSL  
 15 VLLWPERNGVRGNTEIQRLFNLSMN AIRQELETNHAPLKGDVTVLDTLLNNIPNFRDI  
 SILHMESLSKFKLQHPNVVFPALYKELFSIDSQQDLT

## 6. SEQ ID NO: 6 Accession No. NM\_176123 *Drosophila melanogaster*

### 20 Hormone receptor-like in 46 CG33183-PA

1 gaattcattc aactgcaaag agcagccaaa ttgcgcatac gccgcgtatg gccgtcgggtg  
 61 tgagtgcggc tgttcacag cgggtgcac aactgatacc aagtgtacat aactacagct  
 121 acaattgcaa ctatttcacc aatcaacggc agcggcaaca acatcagcaa cagcaccggc  
 25 181 aaacgtttga aacgtcacca aagcttcgca ttcccacta ataattatgt atacgcaacg  
 241 tatgtttgac atgtggagca gcgtcacttc gaaactggaa gcacacgcaa acaatctcgg  
 301 tcaaagcaac gtccaatcgc cggcgggaca aaacaactcc agcgggtcca ttaaagctca  
 361 aattgagata attccatgca aagtctcggc cgacaagtea tccggcgtgc attacggagt  
 421 gatcacctgc gagggctgca agggattctt tcgaagatcg cagagctccg tggtaacta  
 30 481 ccagtgtccg cgcaacaagc aatgtgtggt ggaccgtgtt aatcgcaacc gatgtcaata  
 541 ttgtagactg caaaagtgc taaaactggg aatgagccgt gatgtgttaa agttcggcag  
 601 gatgtccaag aagcagcgcg agaaggctga ggacgaggta cgcttcacac gggcccagat  
 661 gcgggcacaa agcagcgcgg caccggatag ctccgtatac gacacacaga cgccctcgag  
 721 cagcgaccag ctgcatcaca acaattacaa cagctacagc ggccggtact ccaacaacga  
 35 781 ggtgggctac ggagtcctt acggatactc ggccctcgtg acgccacagc agaccatgca  
 841 gtacgacac tcggcggact acgtggacag caccacctac gagccgcgca gtacaataat  
 901 cgatcccga tttattagtc acgcggatgg cgatatcaac gatgtgctga tcaagacgct  
 961 ggccggaggcg catgccaaca caaataccaa actggaagct gtgcacgaca tgttccgaaa  
 1021 gcagccggat gtgtcgcgca ttctctacta caagaatctg ggccaagagg aactctggct  
 40 1081 ggactgcgcc gagaagctta cacaatgat acagaacata atcgaatttg ctaagctcat  
 1141 accgggattc atgcgcctaa gtcaggacga tcagatatta ctgctgaaga cgggctcctt  
 1201 tgagctggcg attgttcgca tgtccagact gttgatctc tcacagaacg cggttctcta  
 1261 cggcgacgtg atgtgcccc aggagggcgt ctacacatcc gactcggaag agatgcgtct  
 1321 ggtgtcgcgc atcttccaaa cggccaagtc gatagccgaa ctcaaactga ctgaaaccga  
 45 1381 actggcgctg taccagagct tagtgctgct ctggccagaa cgcaatggag tgcgtggtaa  
 1441 tacggaaata cagaggcttt tcaatctgag catgaatgcg atccggcagg agctggaaac  
 1501 gaatcatgcg ccgctcaagg gcgatgtcac cgtgctggac aactgctga acaatatacc  
 1561 caatttccgc gatattcca tcttgacat ggaatcgtg agcaagtica agctgcagca  
 1621 cccgaatgtc gttttccgg cgctgtacaa ggagctgttc tcgatatatt cgcagcagga  
 50 1681 cctgacataa caagagcagc agccgttctt ggagacgacc gcggacgatg ttgccgagga  
 1741 tgcggctgcc gccggatgtg tctgccgcc ggtggcgccc cctgccgggc agcaaccagc  
 1801 gctgctcgag gactgagggc cgcaggatgt ggcaacaata attattgag taaacactgc  
 1861 actgcgcatg cagcagatac aagaacttta tcatgattta agctagcata caaccaagga  
 1921 tgtgatctc gccaaaggact cacttaaaaa gaactctatc tatatacata tatatattat  
 55 1981 atatgacaga gcggatgacg caaagggaag ggaaaatatt tcaaaaatat tgttaactca

2041 gtttaagactt ttgcttcgta gagaaccgaa accgaaaccg attgcatttc gagcaagggg  
 2101 catcaaactg attttcgagg ttatactata catatataca cacaacaca cacacacaca  
 2161 tatatatata tgtaacttcc aaactttcat atcctggccc gagcagatca gatcgtctaa  
 2221 gtacttaaaa ccaagcgaaa ttctctacac cgcacaaccc aggaccgta gaccccaata  
 5 2281 attcagttcg gttagtgtta accccagaaa gcccgattcc gatcccgctt aggttgtctt  
 2341 tgccttacgt tgtaactaaa gtatgtgtat tatatatata gcaaatgtat gtataactat  
 2401 gtcgtatcgg ttatatgcct aacaacatta tttttgtaa acaacaaaat cgaatatctc  
 2461 ggaaaatgtg ttcttataat tatattgatt aatgcaatta caatatattt acaatttacc  
 2521 gttacgtttt tacattatac ataagacgca agagaaggaa acggaagttt aaggattaga  
 10 2581 aagctgaata agaaaaggct taaggacgag ctgagtagca gttaaagtga gcgagaaatc  
 2641 gaatgaatac cagaaaattt caagcaagca cataaaagta tgcaatatit tgttataaaa  
 2701 caacttttta ttagtttctt aaatataaca taattacgta catacacaca cgtatatata  
 2761 gggctatata tatctatata tatatatata tacatgatag acaaatccca atccggttcc  
 2821 aaggtttagt aaaaataaag agaaataaaa cgaataacaa aaacttttga tatgaaatcc  
 15 2881 tacgcataat taacaacttt tattgtttct aagacttaaa cttaattaaa atggaaacca  
 2941 aaacagactg accgaccgac cccgacagca tgccacgccc tccccgccc caccctccac  
 3001 agatcctggc agaaatttca aaggagtttg atacacaaat cgagaaaaga aattttcaaa  
 3061 aaaataatat aaagacaagc aaacggcgac tttttggtt gatacatttg aaaagaatat  
 3121 acaattaaat atctgactga ctatacaaag acgttacaca cagcataca catacacaca  
 20 3181 catacacgca tacacacaca gcttacgata cataaattag ttaacttag agtaaacaaa  
 3241 caacaacaaa cacattggat agtaggtgat aattggtgtg tcttaataaa accttaaccc  
 3301 ctccccgacc cccgcccact tgcttaatac ccaacgcccc aaaaagcccc acatttctac  
 3361 taaatgaaaa gcttaataca aactttttg aaattattca agtgaaaatt tcagcaggca  
 3421 ggcataaata ttaattaaca ttaattatag caaggaaact tataaataaa atgtatacaa  
 25 3481 caaaactaca aaaattaaat aaattacatt ttgcaaatc cacaataaat aaacatgat  
 3541 tttgcaaat cacttaaat cctttccctg aatccaagca aaaatattta cactagctta  
 3601 catagaactg ggacgaggac atgaatattt caattgagaa aaaaatctat gttaatgtaa  
 3661 tcgatcgatt tggacatatt taagttcgac atttttggcc ttacaaaaca aaaaacaaaa  
 3721 agaagaaacc taaagtactt tatatatata caaacatat atacaatata gagaatacaa  
 30 3781 aactagtttt aattatata aagcaaggga gcagctttca aactcaaac aaaaatatcc  
 3841 ccgaaaaaaa caacaacttt gttaaaaact gcgcataata aagaaaataa taaacaaagt  
 3901 taatctataa tataaattga agttaagttg atttgagcgg tcgacaacaa gaacataaat  
 3961 gtatctttta atgatatatg tattgttaaa tttgtatgct aagtttttag aaaggttaca  
 4021 ttttaaaga ataataacaa aagatcgaga actcgacaag gtgtaaatg agtacattta  
 35 4081 aattaaaatt tagcatatat aatgcataaa tattatgtta cgatatttac atttatataa  
 4141 aacaaaacaa aaacactaaa gaaaaccgaa aaacagaaag tcccatatta aaatgaaat  
 4201 aaatgagca gaacctataa actgataagg gaattctgaa tatataaaaa aaaaagaaaa  
 4261 ca

40

## 7. SEQ ID NO: 7 Accession No. NM\_079769 Drosophila melanogaster

### Hormone receptor-like in 96 CG11783-PA

MSPPKNCAVCGDKALGYNFNAVTCESCKAFFRRNALAKKQFTCP  
 45 FNQNCDITVVTRRFCQKCLRKCLDIGMKSENIMSEEDKLIKRRKIETNRAKRRLMEN  
 GTDACDADGGEERDHPADSSSSNLDHYSQSQDSQSCGSADSGANGCSGRQASSPGT  
 QVNPLQMTAEKIVDQIVSDPDRASQAINRLMRTQKEAISVMEKVISSQKDALRLVSHL  
 IDYPGDALKIISKFMNSPFNALTVFTKFMSSPTDGVIEISKIVDSPADVVEFMQNLMH  
 SPEDAIDIMNKFMTNPAEALRILNRILSGGGANAAQQTADRKPLLDKEPAVKPAAPAE  
 50 RADTVIQSMLGNSPPISPHDAAVDLQYHSPGVGEQPSTSSSHPLPYIANSPDFDLKTF  
 MQTNYNDEPSLDSDFSINSIESVLSEVIRIEYQAFNSIQQAASRVKEEMSYGTQSTYG  
 GCNSAANNSQPHLQQPICAPSTQQLDRELNEAEQMKLRELRLASEALYDPVDEDLSAL  
 MMGDDRIKPDTRHNPKLLQLNLTAVAIKRLIKMAKKITAFRDMCQEDQVALLKGGC  
 TEMMIMRSVMYDDDDRAAWKVPHTKENMGNIRTDLLKFAEGNIYEEHQKFITTFDEKW  
 55 RMDENIILIMCAIVLFTSARSRVIHKDVIRLEQNSYYYLLRRYLESVYSGCEARNAFI

**8. SEQ ID NO: 8 Accession No. NM\_079769 Drosophila melanogaster**  
**Hormone receptor-like in 96 CG11783-PA**

5  
1 gttattggga ttggcctgga gcaactcggac ggacagtaat tcattaaaat atgtggtgat  
61 aacgcgagct gccgaatctg cgtgcaattc gtgcgtttga cgtgggtact aactgctatg  
121 ctgtcgcgcg gacagttgtt ctgatacgca gatttctgc ctcaccacac acgaccacct  
181 ccattaaaac cagccacccc cccagcgcc tctccaccg acagcagctg ctccaccgca  
10 241 ccaccaggag aggggcaatt aaaaaatcaa tcagagggcc ctaattgaaa gctgccaccg  
301 tcgaaatgtc gccgccgaag aactgcgcgg tgtgcgggga caaggctctg ggctacaact  
361 tcaatgcggg cactgcgag agctgcaagg cgttcttcg acggaacgcg ctggccaaga  
421 agcagttcac ctgccccttc aaccaaact gcgacatcac tgtggtaact cgacgttct  
481 gccagaaatg ccgcctgcgc aagtgcctgg atatcgggat gaagagtga aacattatgt  
15 541 ccaggaggga caagctgac aagcggcgca agatcgagac caaccgggccc aagcgacgcc  
601 tcatggagaa cggcacggat gcgtgcgacg ccgatggcgg cgaggaaagg gatcacaag  
661 cgccggcgga tagcagcagc agcaaccttg accactactc ggggtcacag gactcgcaga  
721 gctgcggctc ggcggacagc ggggccaatg ggtgctccgg cagacaggcc agttgcggg  
781 gcacacaggt caatccgtt cagatgacgg ccgagaagat agtcgaccag atcgtatccg  
20 841 acccgatcg agcctcgcag gccatcaacc ggttgatgcg cagcgagaaa gaggctatat  
901 cggatgatga gaaggtaac agtcacaaa aggacgcctt aaggctggtg tcgatttga  
961 tcgactatcc aggcgacgca ctcaagatca ttcaaagt tatgaactcg cctttaacg  
1021 cgctgacagt attcacaaa tcatgagct caccacgga cggcgttgaa attatctcaa  
1081 agatagttga ttgccccgcg gacgtggtgg agttcatgca gaacttgatg cactcgccag  
25 1141 aggacgcat cgatataatg aacaagtta tgaatacccc agcggaggcg ctgcgcattc  
1201 ttaaccgaat ctaagcggc ggaggagcga acgcagccca gcagacagca gaccgcaagc  
1261 cattgctgga caaggagccg gcggtgaagc ctgcagcgcc agcggagcga gctgatactg  
1321 tcattcaaag catgctgggc aacagtccgc caatttcgcc acatgatgt gccgtggatc  
1381 tgcagtacca ctgcccggg gtgcggggagc agcccagtag atcgagtagc cacccttgc  
30 1441 ctacatagc caactcgccg gacttcgac tgaagacct catgcagacc aactacaacg  
1501 acgagccag tctggacagt gattttagca ttaactcaat cgaatcggtg ctatccgagg  
1561 tgatccgcat tgagtaccag gccttcaata gcatacaaca agcggcatcg cgcgtaaagg  
1621 aggagatgtc ctacggcact cagtctacgt acggtggatg caattcggct gcaacaata  
1681 gccagccgca cctgcagcaa cccatctgcg ccccatccac ccagcagttg gatcgcgagc  
35 1741 taaacgaggc ggagcaaatg aagctgcggg agctgcgact ggccagcgag gctctttatg  
1801 atcccgtgga cgaggacctc agcgcctga tgatgggca tgatgcatt aagcccagc  
1861 acactcgcca caaccaaag ctattgcagc tgatcaatct gacggcggtg gccatcaagc  
1921 ggcttatcaa aatggccaag aagattacag cattccgtga catgtgccag gaggaccagg  
1981 tggccctact caaagggtgc tgcacagaaa tgatgataat gcgtccgta atgattacg  
40 2041 acgacgatcg cgccgcctgg aaggtacccc ataccaaaga gaacatgggc aacatacgca  
2101 ctgacctgct caagtttgc gaaggcaata tctacgagga gcacaaaag tcatcacia  
2161 cgtttgacga gaagtggcgc atggacgaga acataatcct gatcatgtgt gccattgtcc  
2221 ttttacctc ggctcgatcg cgagtgtac acaaagacgt gattagattg gaacagaatt  
2281 cctactatta tcttctgca agatatctgg agagtgtta ttctggctgt gaggcgagaa  
45 2341 acgcgtttat caagctaac caaaagattt cagatgtgga gcgtctgaac aagttcataa  
2401 ttaatgtcta ttgaatgtt aaccatccc aggtggagcc cttgctgcgt gaaatattcg  
2461 attgaaaaa tctactagaca accgatgcgt gtcgggcatt taatgcctat gttgatgccc  
2521 aatgatgaat ggtcaacaag ctgtagttgt tgtgttgtt gatgtctgtt ttatcttgc  
2581 gcttgtaatg ttagatttta atcgaatgtg attgttagat ttgcataac tgcatagatt  
50 2641 ttatattct acatcaaaga gagcatattt aggataccaa gtgcaaagca acacaatcta  
2701 tatgtaatgt acaccgttta ctagtttca aataaactag acgataatgc aataactaac  
2761 ttggaagcgt gggttctgtg caaaaaggaa aaaagacaaa aaaaataaac tgactttgag  
2821 aaccagtgg aa

**9. SEQ ID NO: 9 Accession No. NM\_057539 Drosophila melanogaster****Hepatocyte nuclear factor 4 CG9310-PA**

MMKHPQDLSVTDDQQLMKVKNKVEKMEQELHDPESESHIMHADAL  
 5 ASAYPAASQPHSPIGLALSPNGGGLGLSNSSNQSSNFALCNGNGNAGSAGGGSASSG  
 SNNNNSMFSPNNNLSGSGSGTNSSQQQLQQQQQQQSPTVCAICGDRATGKHYGASSCD  
 GCKGFFRRSVRKNHQYTCRFARNCVVDKDKRNQCRYCRLRKCFKAGMKKEAVQNERDR  
 ISCRRTSNDPDPGNGLSVISLVKAENESRQSKAGAAMEPNINEDLSNKQFASINDVC  
 ESMKQQLLTLVEWAKQIPAFNELQLDDQVALLRAHAGEHLLLGLSRRSMHLKDVLLLS  
 10 NNCVITRHCPDPLVSPNLDISRIGARIIDELVTVMKDVGIDDTEFACIKALVFFDPNA  
 KGLNEPHRIKSLRHQILNNLEDYISDRQYESRGRFGEILLILPVLQSITWQMIEQIQF  
 AKIFGVAHIDSLQEMLLGGELADNPLPLSPPNQSNQDYQSPHTGNMEGGNQVNSSL  
 SLATSGGPGSHSLDLEVQHIQALIEANSADDSFRAYAASTAAAAAAVSSSSSAPASV  
 APASISPLNSPKSQHQHQQHATHQQQQESSYLDMPVKHYNGSRSGPLPTQHSPQRMH  
 15 PYQRAVASPVEVSSGGGGLGLRNPADITLNEYNRSEGSSAEELLRRTPLKIRAPEMLT  
 APAGYGTEPCRMTLKQEPETGY

**10. SEQ ID NO: 10 Accession No. NM\_057539 Drosophila melanogaster****Hepatocyte nuclear factor 4 CG9310-PA**

1 agtgaattc cagtgcgtt ggaagaaaca actgcaaaag gcaaaaacaa agacaatgtt  
 61 tataagctgt atattccgtt ttgattgata taaatgaata tatgcagtgc gccagttata  
 121 caactgccct gcaaaagtca ctcaataaat aaaaaacgcc cgagatgaat ttcacagcgg  
 25 181 cggcaacaag tgcaataata gtaaaaaatc aaaagccaaa caacgaaatc tctcccaaaa  
 241 aaacgaagaa gcgtgtcgcg gtgccaaaaa gaaaacaaaa atagaaaaat acacaacaaa  
 301 ataatacggg gaaacgttaa ttataacgag ccacaaaaat gcataaagaa atcaacaagt  
 361 gtgtgtctgc cttttttcc atattcgtt tcattcatgc ggtcaactca acaataacaa  
 421 ctcaaatag caacaacaac aataacaata tcaacaagag cagcagcagt cgctgataaa  
 30 481 agccctgcag ctaaaacaac acaaaaacaa caaagatagt tagaaagaac atcgtctggc  
 541 cattgagctt taattgccg tcattacttc attactatgt gattggatct tcccgacca  
 601 cttgtaaata aaaagtaaaa atactggta tgaagcatga tgaagcatcc gcaggatctg  
 661 agtgtcacgg atgaccagca gtaaatgaag gtgaacaagg tggagaagat ggagcaggag  
 721 ttgcacgacc ccgaatcgga gagccacata atgcacgcgg atgccctggc ctctgcctat  
 35 781 ccggctgcct cgcagcccca cagtccgatc ggcctcgccc tcagcccca tgccggtggg  
 841 ctgggactga gcaacagtag caaccagagc agcgagaact ttgcgctctg caacggaaac  
 901 ggaaatgcgg gcagcgcagg aggcggaagt gccagcagtg gcagcaacaa caacaacagc  
 961 atgttctcac ccaacaacaa cttgagcgga agcggaagt ggactaacag cagtcagcag  
 1021 caattgcagc agcaacaaca acagcaatca ccgacggctt gcgccatttg tggagatcgg  
 40 1081 gcgacgggca aacattatgg agcctccagc tgcgacggct gcaaggatt cttcaggagg  
 1141 agtgtcagga aaaatcatca gtacacttgc agatttgcg gaaactgcgt tgtggacaag  
 1201 gacaaacgga atcagtgcg ctactgccgg ctgaggaagt gottcaaggc gggcatgaag  
 1261 aaggaggcgg tgcaaaacga gcgggatcgc attagctgcc gccgcacctc caatgacgac  
 1321 ccgatccgg gcaatgggct gtctgtgatt tcttggtta aggcggagaa tgagtgcgt  
 45 1381 cagtgaagg caggcgtgc catggagcca aacattaacg aggaccttc caacaagcag  
 1441 ttgcgagca tcaacgatgt ctgcgagtc atgaagcagc agctgctgac cctggtggaa  
 1501 tgggctaagc agattccggc cttaacgag ctgcagctgg atgaccaggt ggcactgcta  
 1561 cgcgccccatg ctggcgagca ttgctctc ggctgtctc gtcgttcgat gcactgaag  
 1621 gatgttctc tctgagcaa caattgtgtg atcacaaggc actgtccaga tcccttgtg  
 50 1681 tcgccgaatt tggacatctc ccggtcggc gcccgatca tcgatgaact ggtgacggtc  
 1741 atgaaggatg tgggtatcga tgacactgaa ttcgcttga tcaaggccct agtctcttc  
 1801 gatccaatg ccaagggtct taatgaaccg catcgcatca aatcgctacg gcacagata  
 1861 ctcaataatc tcgaggacta catatcagat cggcaatacg agtcgcgcgg tcgcttggc  
 1921 gagattctgc tcatcctgcc ggttctgcag tctattacct ggcagatgat cgagcagatc  
 55 1981 cagtttgcca agatcttgg agtggccac attgattcat tactgcagga aatgtgttg

2041 ggaggagagt tggccgacaa tcctctgccg ctatcgccgc ccaatcagtc aaatgactac  
 2101 cagagtccca cccacacagg caacatggag ggcggtaatc aagttaactc ctctctggac  
 2161 tcgctggcca cgtccggtgg tcctggctcg catagtctgg acctggaggt gcagcacatt  
 2221 caggctctta tcgaggcgaa cagtgcggat gattccttc gggcctacgc ggccagcact  
 5 2281 gcagcggcag ccgctgcage cgtctcgtcc tcctcctctg caccgcacgc cgttgctcca  
 2341 gcctcgatct ctctccgct caacagcccc aagtcacaac atcaacatca gcaacatgcg  
 2401 acgcatcage aacaacagga gagtcctac ttggacatgc cgtcaagca ctacaatggc  
 2461 agtcgggtccg gaccgctgcc aacacagcac agtccccaga ggatgcatcc ctaccaaaga  
 2521 gcagtcgctt cgccggtcga agtgccagc gggggcggcg gattgggtct gcgcaatcct  
 10 2581 gccgatatta cgctcaacga gtacaaccgg agcgagggtg gcagtgcgga ggagctgctg  
 2641 cgacgaactc cactgaagat ccgggctccc gagatgctaa ccgcacccgc tggttatgga  
 2701 acggaaccct gtcgcatgac acttaaacag gagccagaga ctgggtacta gaagaataac  
 2761 gaacgggtgca atatgcagtt tgcaatagga cacccttaa gcacacaacc catacacata  
 2821 caggccctct cttgctgtac tccccaccaa gtgctatata gagatgaaat tgaaatgaag  
 15 2881 aacttactta attgttatgc ctgaacat tttgatactt ttattagtc ctaagtaggt  
 2941 attttgaaa ttgttgctta attttaatg ttaacgcag ttgcaatata ttttggagt  
 3001 catatttgc tcaagaagtt tattatatac aattatacta tatatataca ccatttagca  
 3061 tgtactgagt ttgttggtta tttggttacc ttatacttgt gcgtggatca caaacattc  
 3121 atataaggcc atgcaatata ttgtttagg ttaggtgtt gtctagatta tgctgaaagt  
 20 3181 gtaatatata ttaattta aacaaagaac tattttata tgaatatga taatatacaa  
 3241 actatttc

# **11. SEQ ID NO: 11 Accession No. NM\_176065 Drosophila melanogaster**

## **Hormone receptor-like in 38 CG1864-PC**

25 MDEDCFPPLSGGWSASPPAPSQLQQHLTLQSQAQMSHPNSSNNS  
 SNNAGNSHNSGGYNYHGHFNAINASANLSPSSSASSLYEYNGVSAADNFYGQQQQQQ  
 QQSYQQHNYNSHNGERYSLPTFTISELAAATAAVEAAAAATVSSPSVGGPPPVRAS  
 LPVQRTVSPAGSTAQSPKLAKITLNQRHSHAHALQLNSAPNSAASSPASADLQAGR  
 30 LLQAPSQLCAVCGDTAACQHYGVRTCEGCKGFFKRTVQKGSKYVCLADKNCPVDKRRR  
 NRCQFCRFQKCLVVGVMKEVVRTDSLKGRRGRLPSKPKSPQESPPSPISLITALVRS  
 HVDTPDPSCLDYSHYEEQSMSEADKVQQFYQLLTSSVDVIKQFAEKIPGYFDLLPED  
 QELLFQSASLELFLRLAYRARIDDTKLIFCNGTVLHRTQCLRSFGEWLNDIMEFSRS  
 LHNLEIDISAFACLCALTLITERHGLREPCKVEQLQMKIIGSLRDHVTYNAEAQKKQH  
 35 YFSRLLGKLPELRSLSVQGLQRIFYLKLEDLVPAPALIENMFVTTLPF

# **12. SEQ ID NO: 12 Accession No. NM\_176065 Drosophila melanogaster**

## **Hormone receptor-like in 38 CG1864-PC**

40 1 ctgcccatt ggaggcccc tgcctgtgg cagcagctg cccagcttc aggagaccta  
 61 ctcttgaag tacaacagca gcagcggtag cagccccag caggcgtcct cctcctccac  
 121 cgccgcccc acgcccactg accaggtgct gacctcaag atggacgagg actgctccc  
 181 gcctctgtcc ggccgctgga gtgccagtc gcccgcccc tcccagctcc agcagctgca  
 45 241 caccctgcag tctcaggccc agatgtcga tccaacagc agcaacaaca gcagcaacaa  
 301 cgccggcaac agccacaaca acagtggggg ctacaactac cacggccact tcaatgccat  
 361 caatgccage gccaatctgt cgcccagtc ctcggccagt tcctctacg aatataatgg  
 421 tgtttccgca gcggacaact tctacggaca acagcagcag cagcaacagc aaagctatca  
 481 gcaacataac tacaactcgc acaatggcga gcgttactcg ctgccacgt tcccacgat  
 50 541 ttcggagctg gctgcggcca ctgtgtgtg cgaagctgc gcggcgcca cagtctctc  
 601 ccttcggtg ggccgtccgc cgccagtac cgagcatcg ctgccgttc agcgaaccgt  
 661 ttcgccagcc ggctccacgg cgcagagccc caagctggcc aagatcacac tgaaccagcg  
 721 gcactcccat gcccatgcc atgccctaca gctcaactcg gcaccaatt cgccggcaag  
 781 ttcgccagcg agtgccgcat tgcaggcggg ccgtttgctc caggctccgt cgcagctgtg  
 55 841 tgccgttgt ggccacaccg ccgctgcca gcattatgga gtgcgaacct gcaggggatg

901 caagggattc ttcaagcgga ccgtgcagaa gggctccaag tatgtctgcc tagcggacaa  
 961 gaattgccc gtggacaaga ggcgccgcaa ccgttgccag ttctgccggt tccagaagtg  
 1021 cctggctgta ggcatggtca aggaagtggg gcgcacggac tcgttgaagg gtcgccgagg  
 1081 gagactgccc tcaaaaccga aatgccecca ggagtcgcca ccatcaccac ccatctcgtt  
 5 1141 gatcacggcc ctggttcgca gccatgtcga cagcactccg galccctcgt gcctggacta  
 1201 cagccactat gaggagcagt cgtatgagcga ggcagataag gtgcaacagt ttaccagct  
 1261 gctgaccagc tccgtggacg tgatcaagca gtgcgccgag aagattcccg gctacttga  
 1321 tctcctgccg gaggatcagg agctgtcttt ccagagcgca tcgttgaac tgttcgtcct  
 1381 gcggctggcc tatcgccga ggatcgatga caccaagctg atcttctgca acggcacggg  
 10 1441 gtccaccgc acccagtgc tgcgtcctt cggcgagtgg ctcaacgaca tcatggagt  
 1501 cagccgcagc ctgcacaacc tggagatga catctccgc ttcgctgcc tctgtgccct  
 1561 aacctgac acagaacgcc atggcctgc ggagccgaag aaggtggagc agctccagat  
 1621 gaagatcatt ggcagtctgc gcgaccacgt cactacaat gccgagggcc agaagaagca  
 1681 gcactactc agccgctgc tgggcaagct gccggagctg aggtccctga gtgtccaggg  
 15 1741 actgcagagg atcttctacc tgaagctgga ggacctggg cccgcgccag ctctcatga  
 1801 gaacatgtc gtcaccacat tgcccttcta gaggcgatca tcaagctat catcacaact  
 1861 tgcttctta aactagcccc taagtattgc ctctaggat atacagagaa aggacccat  
 1921 aggacggacg caactagctt tagtagaacc ctgaaataaa taaatctac aacagcaaaa  
 1981 acaaaaccga accgaacaga aatgaagcga atagcagacc caggccatat cttagtgta  
 20 2041 gagctagga gtagccgga cagccccggc tcctcgata attacggaca tgcatattg  
 2101 agaggggggt tccagtgac agcctatggc tcctgcgtga ctctcagca ccgcgagctc  
 2161 caacttggt acgttaattg taaattgt taattcaac tgtcaaaacc ggaatcaacg  
 2221 gccgggcacg caatggcaac actttctat cccggacttc gaagcctgct caacattcgg  
 2281 cactacggac ggacaaaca cggacagaaa cagaactac tctgtctc tgcctttg  
 25 2341 ctaacttcta gtcaattgat ttaggcgaat caaataaata aataaataaa ataagggcgt  
 2401 gcagcagtag tgttatataa ttctatgcc agaccccagc ggttctctc aaggaaatcc  
 2461 ccaatgagt tgcacaaatt gggataaagt acgatagcct attattctta tatttcttt  
 2521 aaaagctga agatagatga gaactgtgt gaaatccact atcatatcat atagttgcta  
 2581 taagccgtgc ttgccctaag ctaagttaga cccgcataaa gttgatagcc caaccaagta  
 30 2641 ttctggttat ttctagact aaggctctaa tagttatagg ctaagactat tctgttcgat  
 2701 ttatcaatgc accaaacagt gcacaatgag agtataagta ccttctgtg atgattgtgt  
 2761 ctgacacaga gagagttgca cacaagcaca caaactagcc gataagttac taaatacgt  
 2821 ctaatatcta atatataata tataatataa tatatataag tccaagtatt cggaaatcca  
 2881 agaacccttg cataaccgca gtctgtacgt tccaaacgag aaaagaactt tatttaatcc  
 35 2941 tagaccactc catctaagtt ctcaaagaat cgtatgtgga tcgttggatc tgtctctcta  
 3001 tatatgtgtg tgtgttatct cgtatagaaa cccctctatg tgattttgtg atagattggc  
 3061 attgaactct atatatttat atatatatgt ctataatata tatacacgca taaatatata  
 3121 ttttatgtc taactttgt atggtttatt ttatagctac cactttctt tgataacaaa  
 3181 aagtaaaaaa ctcttagat agcaaatatt tcaaaggat gtacgagga ctttcaaag  
 40 3241 taccagtct tagcgactt ccaattaacg ttctattaa cgaaagacag atttctatg  
 3301 tgttaaattg aagacttcta taactataac taaatgcaag ctaagagcaa aaacacaaat  
 3361 ccacaaatcc ccaaagtga taacatatct ctcaagctt tcgagtgcac ggaacacgta  
 3421 gaaccgaaac ccaagtgtta ctaaatccat ttaataatcg gcaagccggg ggcgtcggcg  
 3481 tggtaatac gttctatta cctatacaat ttatagat catattaaa ttattgtaca  
 45 3541 ttagcacat gaaatgtcg acaactagat ttgtaccat cttaaagaag aacctaggcc  
 3601 aagctaaact aagtataaac tatgatctgc atgcggctga gctgtagcta tgagaaatat  
 3661 acctgcgtgg atctaagtga aatgggacac ttgaattta gatatgaaac gttctaaacg  
 3721 cgacgtacta actctccaa ctgcgaact taccaattaa gagaaattcc cagaaaatgt  
 3781 gtcaggattt caaagcgtc catctactt gaaccaccc aatcaacaaa tacaatcct  
 50 3841 agggaagtg agaggtcag caaccataga gcaatattc ataagaaaac gcacctaaa  
 3901 ttaccgaaa acatagatta acctgatct gtaacgttg ggagcgataa taagccagga  
 3961 ttaaacagga acagttagg gaccaatca gtgcgaaac agatgataga taggttcggg  
 4021 ttcgaaacc taaacgcgt gccatttag ccgttacaac attggatc aacctgcac  
 4081 atgaatatga atatgaatat gaattatata gagatatat tagctatagg aacctactt  
 55 4141 gtacctacac gacatggaaa catcaaacct acatgcata ttacacacat atatttgaa  
 4201 tagagcgacg acttttaca gttgcgtaca aagctatagc tatagctga tatggccatc  
 4261 ccagagcgag catatacata tttttgggt tattgtctt ttgtaattt ataaatgcat  
 4321 acatatttat tgtactacgt gaatgtcaag tgtggattca tttttgag atacagctac

4381 aaaacgaaac aaaagaaaat aaaacaaaac agaagagtaa acgtgaaatt ttctgatgaa  
 4441 acaattttaa atgagaactt ttaatatgt ctattaaagg atatacatat acacactaac  
 4501 atacatatat attttactat gtaacggata gaattaagct agatgcagcg cataaagctt  
 4561 tataacaaca attgaaaagc aacagaagaa attggcacia attaaattta tatagcataa  
 5 4621 ttagacgtcc ttgcaagat aatgttattc gtaataagag cgtcaatcgg tacatcgggc  
 4681 gctatttccc actacacccc caaccacaca atagataacc taagctatgt atgtacatta  
 4741 gctatgtata tccagccccc ttatgcgcct actactagaa atgcagaaag cagaaagaga  
 4801 ggtgaaacct atagacgcta tcacaaatgt ctatctgata gacatcggta ctaccaatgc  
 4861 tatattgcca gttgtgtaat ttactcttat ttgatcgttt catttaccag ttaagaaccc  
 10 4921 aaatcatata agtggtatga tggaagaact ataacttgca attcaattaa ctctgcaata  
 4981 cgataacaag caaagcgaat catttcattt cgatttaate ttaattata tatacttaaa  
 5041 cgatgtaagc ccaaaacaaa cgtttttct atatctgtct tttagcaaaa ttagttatac  
 5101 gcaaaaccaa accgtattta cataaatgta taaaaacaa atcgtatatt ttcatiggtt  
 5161 tgaaataaat acataaaaca a

**13. SEQ ID NO: 13 Accession No. NM\_141390 Drosophila melanogaster  
 CG10296-PA**

20 MSNFSACAVCGDQSSGKHYGVSCCDGCSCFFKRSVRRGSSYACI  
 ALVGNCVVDKARRNWCPSCRFRCLAVGMNAAAVQEERGPRNQVALYRTGRRQAPPS  
 QAAPSPTPHSQALHFQILAQILVTCLRQAKANEQFALLDRCQQDAIFQVWSEIFVLR  
 ASHWSLDISAMIDGCGDEQLKRLICEAHQLRADVLELNFMESLILCRKELAINAEYAV  
 ILGSHSKAALISLARYTLQQSNYLRFGQLLLGLRQLCLRRFDCALSCMFRSVVRDILK  
 25 TL

**14. SEQ ID NO: 14 Accession No. NM\_141390 Drosophila melanogaster  
 CG10296-PA**

30 1 atgtcgaact tcagtgcctg cgcagtggtc ggcatcaga gctccgggaa gcactacggc  
 61 gtgtcctgct gcgatgggtg ctctctgttt ttaagcggga gcgtgcggcg cgggagcagc  
 121 tacgcctgca tcgtctggt cgggaactgt gtggtggaca aggcgcggcg gaactggtgt  
 181 ccctcctgcc gctccagcg atgcctggcc gtgggaatga acgctgctgc ggttcaggag  
 241 gagcgcggtc cgcgcaacca gcagtggtct ctctaccgca ctggccggag acaagctccg  
 35 301 ccatctcagg cggcgccatc cccgacgccc cactcccagg cgtgcactt ccagatcctc  
 361 gccagatcc ttgtcacgtg cctgcgccag gcgaaggcca acgagcagtt cgtctgttg  
 421 gatcgtgcc aacaagacgc catcttcag gtggtgtgga gcgagatctt cgtcctgca  
 481 gcgtccact ggtctctgga catcagcgc atgacgacg gctgcggcga tgagcagtc  
 541 aaacggctca ttgcgagcg ccaccagcta agggccgacg tcttggaact caactttatg  
 40 601 gagtccctaa tctgtgcag aaaagaattg gccatcaatg cggagtatgc cgttatcctg  
 661 ggaagccact ctaaagcgc cctgatctcc ttagcccgt acaccctgca gcaatccaac  
 721 tacctgcggt tcggacaact gctccttggc ctgaggcagc tgtgcctgag gcgcttcgac  
 781 tgcgcgcttt cttgtatgt tcgcagcgtg gtcagggaca tcttaaaaac actttag

**15. SEQ ID NO: 15 Accession No. NM\_169459 Drosophila melanogaster  
 seven up CG11502-PC**

50 MGMRREAVQRGRVPPTQPGLAGMHGQYQIANGDPMGLAGFNGHS  
 YLSSYISLLLRAEPYPTSRYGQCMQPNMIDNICELAAARLLFS AVEWAKNIPFFPE  
 LQVTDQVALLRLVWSELFVLNASQCSMPLHVAPLLAAAGLHASPMADRVVAFMDHIR  
 IFQEVEKLKALHVDSEYSLKAIVLFTTDACGLSDVTHIESLQEKSQLALEEYCRT  
 QYPNQPTRFGKLLRLPSLRVSSQVIEQLFFVRLVGKTPJETLIRDMLLSGNSFSWP

YLPSM

**16. SEQ ID NO: 16 Accession No. NM\_169459 Drosophila melanogaster  
seven up CG11502-PC**

5

1 ctaaattggt gtttcaaaa gaaatgaatt tctttccact cctttcagaa ttcaagaata  
 61 aatattgaag caatatggct tcccttggtc aaaccgatca atcgttgcaa atctttcttc  
 121 aagcgctcgg tgcgacgtaa tctaacttac tcttgccgcg gcagcagaaa ctgtcccata  
 181 gatcaacacc atcgcaatca atgtcaatat tgtcgattga agaagtgcct caaatgggc  
 10 241 atgagacgcg aagctgttca acgtggacgc gtaccacca ctcagcccgg tctggccggc  
 301 atgcatgggc agtaccagat tgccaacggg gatcccatgg gcattgccgg cttaacggg  
 361 cactcgtacc tcagttccta catctcgctc ctgctgcggg cggaaccgta tccgacttcg  
 421 cgatatggcc agtgcattga acccaacaac attatgggca tcgacaacat ctgcgaactg  
 481 gccgcccgcg tgctcttctc ggcggtcgag tgggccaaga acataccctt ctcccggag  
 15 541 ctgcagggtga ccgaccagggt ggccctgctc cggctcgtct ggtagagct ctctgtccta  
 601 aacgccagcc agtgcctcat gccgtccat gtggcgccac tgctggccgc cgccggactt  
 661 catgcctccc cgatggccgc cgatcgtgtg gtggccttca tggaccacat ccgcattctc  
 721 caggagcagg tggagaagct gaaggcgctg catgtcgact ccgaggagta ctctgcctc  
 781 aaggcgatcg tgctcttctc caccgatgcc tgcggcctgt ccgatgtgac gcacattgaa  
 20 841 tccctgcaag agaagtcgca gtgcgccctc gaggaatact gccggacca gtatcccaac  
 901 cagccacga gattcggcaa gctgcttctc agactgccat cgtgcgaac ggtctcctca  
 961 caagtcattg agcaattgtt tttgtgcgt ctatcgga aaacgccaat tgaaacgctg  
 1021 atacgcgata tgctgctgag cggcaacagt ttctcctggc cctatctgcc ttgatgtga  
 1081 cacacgatgt ggcgccaatt gacaacaact tgatcatcgg ccgcagctgt ggcggctgca  
 25 1141 acgctcaaca tcaattccgg cggaggcggc atcggcatcg gcggcggggg cagtggcagt  
 1201 ggcgggtggcg gtagtggagg cgtggtggga gtcgttgat gtggcagcca caacgttgc  
 1261 gctgccagtc atgaccagct cgcaatgtt gctgtcatgc agcaaacata cggcagcggc  
 1321 ggcagcagca gcagcagcat cagcgttgc cacaacgga acaacggcag cggcggcagc  
 1381 atttgaatc agcagatcaa caactacggc aacaacagca acaacaatgt cggcaatcat  
 30 1441 atgagtgcag gcagttttt cgggtgggtcc aacaacagca tccacagtag tggcaatagc  
 1501 aataccgatt atatgaccac gccagccacc gcttatgca caccagcagc agcagccaca  
 1561 tccacggtga acaccacaac gatgctgtct aattactgcg atgccgccac catgatgatg  
 1621 gccgctgctg cagtcaatgc aatcaatgc ctgcagcaac atcaccagcg catgttgcct  
 1681 gcgggcagca gcaacagcag cagcaacaac agcagcagca acagcaacgg cgcagcagca  
 35 1741 atgccctct catctcgtc tggctcactg tcctctgct catcgacccc aacagcaaca  
 1801 gcaactgcga ctgcaattgc aacagcaaca gcaactgcag cagcaacagc cgcgagcaaa  
 1861 caacagcaac aatgcgcccc aaatttaac gatatcagcg aagtctctct cattgtggat  
 1921 gtcaagtagt gtaattattt atgcatctag aaatggggct ataaaccaac ctgttagata  
 1981 cccgcccccg ccccaccac taccacaaa accataaaac cccaaaaaaa aaacaattga  
 40 2041 aaaatgtaaa aaaaaaaagt tggaggatga gcgcgcgla gcttaattga ctaatttcc  
 2101 attttagct tttgtgtaa cttgtacat aactctcga aaaattcaag ttttctcta  
 2161 ggccacccca gctgtgagca aaaccaatct cagctgacat atccaagaga acttcaaaag  
 2221 tgaagcccc aaaaaaagta agaaggcgcc aaaaaaacgt cttacatat gaatgtgtat  
 2281 aatatttaa tggcactgag ttctacttaa tttagacca caaacactg aaaaaatcaa  
 45 2341 tgaaaaaata agaattgtgg aaagagaaaa atcccccta acactttcaa aagacaaaac  
 2401 ataaagatag ttaaaatatt tatatatgta atgtagcata tacacgtata tagtacatat  
 2461 atgaatatat aaacgaaact ctactcccag tggtttgag aaatatacca aaaattttaa  
 2521 gctatgttta ctgatgtgt ggcaatttt atgtgtgct tagcaatttt attttactt  
 2581 taagtaaaat taaaattta taaacattcg attctcgact ggttttctc ggcggatgta  
 50 2641 tctcaaagat gcttctgtat gggaaggccg aattgttgaa atacgaatgc aaaatttagc  
 2701 gaattttta tttagtaacc attacgagta aaacacaaa atgttcagt caagtttcag  
 2761 ttcttaaacg atttttctg aagcttaagc attatcttat ttatgtgtat agagtatgaa  
 2821 aagtttcta tatttgtaa taataaaaat ttgcgttat aatgaa

**17. SEQ ID NO: 17 Accession No. NM\_079857 Drosophila melanogaster****tailless CG1378-PA (tll) mRNA**

5 MQSSEGSPDMMDQKYNSVRLSPAASSRILYHVPCKVCRDHSSGK  
 HYGIYACDGCAGFFKRSIRRSRQYVCKSQKQGLCVVDKTHRNQCRACRLRKCFEVGMN  
 KDAVQHERGPRNSTLRRHMAMYKDAMMGAGEMPQIPAEILMNTAALTGFPGVPMMPG  
 LPQRAGHHPAHMAAFQPPPSAAAVLDLSVPRVPHHPVHQGHGFFSPTAAYMNALATR  
 ALPPTPPLMAAEHIKETAAEHLFKNVNWIKSVRAFTELPMPDQLLLLEESWKEFFILA  
 10 MAQYLMPMNFAQLLFVYESENANREIMGMVTREVFHAFQEVNLNQLCHLNIDSTEYECLR  
 AISLFRKSPPSASSTEDLANSSILTGSGSPNSSASAESRGLLESGKVAAMHNDARSAL  
 HNYIQRTHPSQPMRFQTLLGVVQLMHKVSSFTIEELFFRKTIGDITIVRLISDMYSQRKI

**18. SEQ ID NO: 18 Accession No. NM\_079857 Drosophila melanogaster****tailless CG1378-PA (tll) mRNA**

15 1 gagtccacat cggagtaacc aaggatatat cgaatatatc acacaatccg caataccgcc  
 61 gtccacccaa accgttaaaa caaaaatcca aaacgactca aagatacacc agtgccaagt  
 121 gaaattcaat ttgtgcaagc gttttacaa aaatcgccaa aattacgccc cacatcggtg  
 181 tgcagtcgtc ggagggttca ccagacatga tggatcagaa atacaacagc gtgcgtcttt  
 20 241 cgccagcggc atcgagtcgc attctatacc atgtgccctg caaagtctgc agagatcaca  
 301 gctccggcaa gcattacggc atctacgcct gtgatggctg cgccggatc ttcaagagga  
 361 gcattcggag atcccggcag tatgtgtgca agtcgcagaa gcagggactc tgtgtggtgg  
 421 acaagacgca caggaaccaa tgtagggctt gccgactgag gaagtgctt gaggtcggaa  
 481 tgaacaagga tgcagtgcag cacgagcggg gaccgcggaa ctccactctg cgtcgccaca  
 25 541 tggccatgta caaggatgcc atgatgggcg ccggcgagat gccacaaata cccgccgaaa  
 601 ttctgatgaa cacggctgcc ttgaccggct ttctggagt accgatgcc atgcctggcc  
 661 tgcccagag ggctggctcat catctgctc acatggctgc ctccagccg ccaccatcgg  
 721 ctgccgtgt cttggacta tccgtgccac gattgcccc taccgggtg caccaaggac  
 781 accacggtt cttctgccc accgccgct acatgaatgc cctggccact cgggccctgc  
 30 841 cccccactcc tccgtgatg gcagctgagc acatcaagga aaccgcggcg gaacacctat  
 901 tcaagaacgt caactggatc aagagcgtac gggccttcac cgaactgcc atgccggatc  
 961 agctgctcct gctggaggag tcctggaagg agttcttcat cctggccatg gccagttacc  
 1021 taatgccc atgaattgcc cagctgctgt tcgtctacga gtccgagaat gccaacggg  
 1081 agatcatggg catggtgacc cgcgaggtgc acgccttcca ggaggtgctg aaccaactgt  
 35 1141 gccatctgaa catgacagc accgagtacg agtgtctgag ggctatttcg ctctccgta  
 1201 agtaccacc gtcggcaagt tctaccgagg atttagccaa cagctcaatc ctgacaggaa  
 1261 gcggcagccc gaactcctc gccctgctg aatccagggg tctctggag tcgggaaaag  
 1321 tggcggccat gcacaacgat gcccgagtg cgctgcacaa ctacatccag aggaccatc  
 1381 cctcgcagcc catgcgattc cagacgctt tgggcgtggt gcagctgatg cacaaggct  
 40 1441 caagcttcac catcgaggag ctgttcttc gaaagacct cggcgacatc accattgtgc  
 1501 gcctcatctc cgacatgtac agtcagcgca agatctgaaa agtatgtaga gcctagacta  
 1561 atcgccgcac tcgaagtgcc ttccaagtgc tgggaactgt gataatctc gaagaagcgc  
 1621 ttggacaat actcgatcag tgaatcaac gatttctcat atccaggagt cgagcctta  
 1681 aatactaca caactcacc ctaatacct tacctaaaca gaactcgaag taatcttagc  
 45 1741 taaagtctc cagaccatcc agatgtgtt caaattgcat tcgcaaaagt ttcaacttg  
 1801 cctgttaaat acgtcaatc tagttttaa cacttagtt ttaagcgc atatttagct  
 1861 ttaggattg gaaaaataat tatic

**19. SEQ ID NO:19 Accession No. NM\_057792 Drosophila melanogaster****50 dissatisfaction CG9019-PA**

MGTAGDRLLDIPCKVCGDRSSGKHGYIYSCDGC SGFFKRSIHRN  
 RIYTCKATGDLKGRCPVDKTHRNQCRACRLAKCFQSAMNKDAVQHERGPRKPKLHPQL

5

15 **dissatisfaction CG9019-PA**

20

25

30

35

40

45

50

55

2401 cgtcgtcata tgcgaactta ttgtattcc aatgcgaccc gaatectatt cagattcact  
 2461 gcggcaggag gcggtccaaa tgtggggcgg aagctgcaga tgctatggt cgcaggacgc  
 2521 catgtaatgg aggcgtatgt actaaccgcg ctctccatt ggcgatgcag tccgcgatga  
 2581 tggegcactc ccacacccac acccgtaacc acacettgat ttatgccgg caatgcgtcg  
 5 2641 gagtctcctt actttcgctt cgttttctaa catttgtatc ctattttat ttatctttt  
 2701 tccacggatt ttctgtttg actgcctggg cggcactctt tatttatctt tcattcgacg  
 2761 tttgtcgtc gcttttctaa aaattcccca tgttatttca acctggcaag gacctcgag  
 2821 tccattccc gcgcccttac ttacaaatca ctteccatcc cacatccagc aattccgtgg  
 2881 ttgaattct tctgtcatt gactacgaaa taccctttaa tcagacaaat aaagaatatt  
 10 2941 agttgtaatt ctttttctg caatccagct ctaaaacggg tttcttaac gaaatcgata  
 3001 aatgtaaaaa ttatacatat cctttacca cattgtttgc cta

## 21. SEQ ID NO: 21 NM\_166092 *Drosophila melanogaster* CG16801-PA

15 MATGRSLLFRVPWYVCLCVCAESAEPGVYWRLRLRLGLPTLAGP  
 HTNTLTLTARTSSCRSIKKERIKASQQANAPPELPLKVSVDVNIILAAHSQRRRIGLV  
 RFHQRESEDRPLAVASPRILQINMEPTAMNPKKLHSPQRHCYTPPPAPMHGQAPPPTST  
 GVAPPTQPPPPHPAAPNVPNGRLLSWNHSAAAAAAAAAAAAAQAANSNMNHSSAAEGSSMT  
 RIKGQNLGLICVVCBDTSSGKHYGILACNGCSGFFKRSVRRKLIYRCQAGTGRCVVDK  
 20 AHRNQCQACRLKKCLQMGMNKDDSDIDVTNDNEEPHAVSRSDSSFIMPQFMSPNLYTH  
 QHETVYETSARLLFMAVKWAKNLPSFARLSFRDQVILLESWSELFLNLIQWCIPLD  
 PTGCALFSVAEHCNNLENNANGDTCITKEELAADVRTLHEIFCKYKAVLVDPAEFACL  
 KAIVLFRPETRGLKDPAQIENLQDQAHHTKTQFTAQIARFGRLLMLPLLRMISSHI  
 25 ESIYFQRTIGNTPMEKVLCDMYKN

## 22. SEQ ID NO: 22 NM\_166092 *Drosophila melanogaster* CG16801-PA

1 atggcgaccg ggcgttctt gctcttcga gtgccttgg atgtgtgctt gtgtgtgtgc  
 30 61 gcagagagcg cagagccggg tgtttattgg agattgcgat tgcggcttgg ctaccacaca  
 121 ctgcgaggcg cgcacaccaa cacactaaca ctaacagcga ggacaagctc ctgccgcagc  
 181 atcaagaagg aacgaatcaa agcaagccaa caagcaaatg cggcaccaga gttgccacta  
 241 aaagtctccg ttgacgttaa catcatcatc gcggcacact cgcagcgccg tcggatcgga  
 301 ttggttcggg ttatcagcg ggaatcagag gaccgtccac ttgccgtcgc ctctccacga  
 35 361 ttgcaaatga atatggagcc tactgcgatg aaccgaaaa aactccacag tccgcagcgg  
 421 cattgctaca ctccgccgcg ggcgccgatg caggacagg cgcctccacc tacatcaacg  
 481 ggcgtggccc cgcacacaca gccaccgcc cctcatccc cgcgccaaa cgtgcccaat  
 541 ggtcgattgc tgagctggaa tcacagtgc gctgcagctg ctgcggcggc ggcagcccaa  
 601 gcggcagcca actccatgaa cactcgtcg gcggcgagg gttcatgat gaccgggatt  
 40 661 aagggtcaga acctgggct catctgcgtg gtgtgcggcg acaccagctc gggaaagcac  
 721 tacggaatcc tagcctgcaa tggctgtcc ggattcttca aacgcagcgt gcggcgaaa  
 781 ctcatctatc gctgccaggc gggaacggga cgtgtgtgg tggacaaagc tcacgggaat  
 841 caatgccagg cctgcaggct caagaagtgc ctcaaatgg gaatgaacaa ggacgacgac  
 901 tccatagatg taaccaacga caacgaggag ccgatgcag tcagcagatc ggattcgagt  
 45 961 ttattatgc cgcagttcat gtcgccaat ctgtacacc atcaacacga aacagttac  
 1021 gagacaagtg cccggctgct ctcatggcc gtcaagtggg ccaagaacct gccagctt  
 1081 gcaagacttt ctttcggga tcaggtaat ttgctggagg agtctggtc ggagctgtc  
 1141 ctgctgaacg caatccaatg gtgcattccc ctggatccca ccggtgcgc cctctctcg  
 1201 gtggcggagc actgcaataa tctagagaac aatgccaatg gcgacactg catacaaaag  
 50 1261 gaggagctgg cggcgatgt gcaacgctc cagagatct tctgcaata caaggcggg  
 1321 ctggtggacc ccgtgaatt cgcgtgcctc aaggcgatag ttctctccg gccggaaacg  
 1381 cgcggactta aagatccgc gcagatagag aatcttcagg atcaggcgca ccacacaaag  
 1441 acgcagttca ccgccagat agccagattc ggacgactcc ttctcatgct gccgttctg  
 1501 cgcagatca gctccacaa gattgagtc atctatttc agcgactat tgggaacag  
 55 1561 cccatggaaa aggtgctctg tgacatgtat aagaactag

**23. SEQ ID NO: 23 Accession No. NM\_168258 Drosophila melanogaster  
estrogen-related receptor CG7404-PA (ERR)**

MSDGVSIHLHIKQEVDTPSASCSPSSKSTATQSGTNGLKSSPSV  
 5 SPERQLCSSTTSLSCDLHNVSLNDGDSLKSGTSGGNGGGGGGGTSGGNATNASAGA  
 GSGSVRDELRLCLVCGDVASGFHYGVASCEACKAFFKRTIQGNIEYTCPANNECEIN  
 KRRRKACQACRFQKCLLMGMLKEGVRLDRVRGGGRQKYRRNPVSNSYQTMQLLYQSNTT  
 SLCDVKILEVLNSYEPDALSVQTPPPQVHTTSITNDEASSSSGSIKLESSVVTPNGTC  
 IFQNNNNNDPNEILSVLSDIYDKELVSVIGWAKQIPGFIDLPLNDQMKLLQVSWAEIL  
 10 TLQLTFRSLPFNGKLCFATDVWMDEHLAKECGYTEFYHCVQIAQRMERISPRREEYY  
 LLKALLLANCDILLDDQSSLRAFRDTILNSLNDVVYLLRHSSAVSHQQQLLLLLPSLR  
 QADDILRRFWRGIARDEVITMKKLFLEMLEPLAR

**24. SEQ ID NO: 24 Accession No. NM\_168258 Drosophila melanogaster  
estrogen-related receptor CG7404-PA (ERR)**

1 ccctggtcag gtctggttca ccaaaaaaga aaataaaatt acatttcaat cttccaata  
 61 tgcaaatatc tgcacgaaaa ccagcgagaa cagcatgctc acaataaaga gcccccaaac  
 20 121 aatgtgactc gtatccgcgc agagtgcgtc ttcgtgcctt gcccagatgc caaatccaaa  
 181 tcccaatcca ggcgcacaaa atcgatgcag atgctgtctg catttcata gaaagtgcac  
 241 ctgaataacc gatggtcgcc aaaagccacg atgtccagta ataatgacca gtgaataaac  
 301 aattatgact cgagcatcga aaaatgctga ggaacgaata cataagcaat aacaagaagg  
 361 tgctcaactc ggacccaaac aagtactaca tgctaacggc cgaggaggcc gatatgtatt  
 25 421 gacgttggtc cagtggagct gattacacaa aagatcctca gaacgatttt atccaaggca  
 481 cgaacatgct cgacggcgtc agcatcttgc acatcaaaca ggagggtggac actccatcgg  
 541 cgtcctgctt tagtcccagc tccaagtcaa cggccacgca gaggggcaca aacggcctga  
 601 aatcctcgcc ctggtttcgc cggaaaggc agctctgcag ctgcagacc tctctatcct  
 661 gcgatttgca caatgtatcc ttaagcaatg atggcgatag tctgaaagga agtggtacaa  
 30 721 gtggcggcaa tggcggagga ggagggtgtg gtacgagtgg tggaaatgcg accaatgcga  
 781 gtgccggagc tggatcgga tccgtcaggc acgagctccg ccgattgtgt ttggtttgtg  
 841 gcgatgtggc cagtggatc cactatggtg tggcgagtgt tgaggcttgc aaagcgttct  
 901 ttaaacgcac catccaaggc aacatcgagt acacgtgtcc ggcgaacaac gagtgtgaga  
 961 ttaacaagcg gagacgcaag gctgccaaag cgtgtcgtt ccagaaatgt ctactaatgg  
 35 1021 gcatgtctaa ggagggtgtg cgcttgatc gattcgtgg aggacggcag aagtaccgaa  
 1081 ggaatcctgt atcaaatct taccagacta tgcagctgct ataccaatcc aacaccact  
 1141 cgctgtgcga tgcaagata ctggagggtc tcaattcata tgagccggat gccttgagcg  
 1201 tccaaacgcc gccgccgcaa gtccacacga ctgcataac taatgatgag gcctcactct  
 1261 cctcgggcag cataaaactg gattccagcg ttgttacgcc caatgggact tgcatttcc  
 40 1321 aaaacaacaa caaatgat cccaatgaga tactaagcgt ccttagtgat atttacgaca  
 1381 aggaatttgt cagcgtcatt ggctgggcca agcagatacc tggctttata gatctgccac  
 1441 ttaacgacca gatgaagctt ctccagggtg cgtgggcaga gatcctgacg ctccagctga  
 1501 ccttccggtc cctaccgttc aatggcaagt tatgtctgc cacggatgtc tggatggatg  
 1561 aacatttggc caaggagtgc ggttacacgg agttctacta ccactgcgtc cagatcgac  
 45 1621 agcgcatgga aagaatacgc ccacgaaggg aggagtacta cttgctaaag gcgtcctgc  
 1681 tggccaactg cgacattctg ctggatgatc agagttcct gcgcgcattt cgtgatacga  
 1741 ttcttaattc tctaaacgat gtggtctact tgetgcgtca ttcgtcggcc gtgtcgcac  
 1801 agcaacaatt gctgctttg ctgccttcgc tgcggcaggc ggaatgatac ctgcgaagat  
 1861 ttggcgtgg aattgcacgc gatgaagtca ttaccatgaa gaaactgtc ctgcagatgc  
 50 1921 tcgagccgct ggccagggtg aaaggattat gcgggcgccc aaactagttg atctagctga  
 1981 taagcaaagg tgcaaatata gtcttaggta tatatggatg tatactagag tagattaagc  
 2041 gtaggataag ccatgtatat aaatagtaa atactgtcg ggtaagatta gtgcgcagaa  
 2101 aaaatctctt ttaatggact accaactaca gcaactggaa aaccctactt atctctaga  
 2161 atcggggtgt gcttactcgt gtaaaaggcg catataggtg ttatgtgtct aaagttgtga  
 55 2221 gtcacagatc tcaataaatt tgttcaattc tactgggtc tgatatatgt atatccgca

2281 acctctgat gtaacgtatg aattgtggg cacttttaa atacgatagt ggttctacaa  
 2341 tacaatggat tatactgttt ctaagtgtca tgtaaccag tgattctgtg tctatgtgt  
 2401 acacatgcgg tcaaaagaat agcaatgtcg tccgtgaata ataaaccgtt tgtaactgt  
 2461 gttccatac tccctaagtt ctgtattctt tggggatttt ctttccctaa acaaattcaa  
 2521 attagtttt

## 25. SEQ ID NO: 25 Accession No. NM\_168908 *Drosophila melanogaster*

### Hormone-receptor-like in 78 CG7199-PC

10 MDGVKVETFIKSEENRAMPLIGGGASGGTPLPGGGVGMGAGAS  
 ATLSVELCLVCGDRASGRHYGAISCEGCKGFFKRSIRKQLGYQCRGAMNCEVTKHHRN  
 RCQFCRLQKCLASGMRSDSVQHERKPIVDRKEGIIAAAGSSSTSGGGNGSSTYLSGKS  
 GYQQGRGKGHSHVKAESAATPPVHSAPATAFNLNENIFPMGLNFAELTQTLMFATQQQQ  
 QQQQQHQQSGSYSPDIPKADPEDDEDDSDMDNSSTLCLQLLANSASNNNSQHLNFNAGE  
 15 VPTALPTTSTMGLIQSSLDMRVIHKGLQILQPIQNQLERNGNLSVKPECDSEAEDSGT  
 EDAVDAELEHMELDFFECGGNRSGGSDFAINEAVFEQDLLTDVQCAFHVQPPTLVHSHYL  
 NIHYVCETGSRIIFLTIHTLRKVPVFEQLEAHTQVKLLRGVWPALMAIALAQCQGQLS  
 VPTIIGQFIQSTRQLADIDKIEPLKISKMANLTRTLHDFVQELQSLDVTDMFGLLRL  
 ILLFNPTLLQQRKERSLRGYVRRVQLYALSSLRRQGGGIGGGEERFNVLVARLLPLSSL  
 20 DAEAMEELFFANLVGQMMDALIPFILMTSNTSGL

## 26. SEQ ID NO: 26 Accession No. NM\_168908 *Drosophila melanogaster*

### Hormone-receptor-like in 78 CG7199-PC

25 1 attggaacaa ggagatttta ttgcgttaga aaaggttcaa aataggcaca aagtcctga  
 61 aaatatcgta actgaccgga agtaacataa cttaaccaa gtgcctcgaa aaatagatgt  
 121 ttttaaaage tcaagaatgg tgataacaga cgtccaataa gaatttcaa agagccaaat  
 181 gtttgggttt cagttattta tacagccgac gactatttt tagccgcctg ctgtggcgac  
 30 241 aatggacggc gtttaaggtt agacgttcat caaaagcgaa gaaaaccgag cgatgccctt  
 301 gatcggagga ggcagtgcct caggcggcac tcctctgcca ggagggcgcg tgggaatggg  
 361 agccggagca tccgcaacgt tgagcgtgga gctgtgtttg gtgtgcgggg accgcgcctc  
 421 cgggcggcac tacggagcca taagctgcga aggctgcaag ggattctca agcgtcgtat  
 481 ccggaagcag ctgggctacc agtgcgcgg ggctatgaac tgcgaggtca ccaagcacca  
 35 541 caggaatcgg tgccagtct gtcgactaca gaagtcctg gccagcggca tgcgaagtga  
 601 ttctgtgcag cacgagagga aaccgattgt ggacaggaag gaggggatca tcgtgctgc  
 661 cggtagctca tccactctg gcggcggtta tggctcgtcc acctacctat ccggcaagtc  
 721 cggctatcag caggggcgtg gcaaggggca cagtgtaaag gccgaatccg cggccacgcc  
 781 tccagtgcac agcgcgccag caacggcctt caatttgaat gagaatatat tccgatggg  
 40 841 ttgaatttc gcagaactaa cgcagacatt gatgttcgt acccaacagc agcagcaaca  
 901 acagcaacag catcaacaga gtggtagcta ttgccagat attccgaagg cagatccga  
 961 ggatgacgag gacgactcaa tggacaacag cagcacgctg tgcttgagtg tgctgcgcaa  
 1021 cagcggcagc aacaacaact cgcagcacct gaactttaat gctggggaag taccaccgc  
 1081 tctgcctacc acctcgacaa tggggcttat tcagagttcg ctggacatgc gggctatcca  
 45 1141 caagggactg cagatcctgc agcccatcca aaaccaactg gagcgaaatg gtaatctgag  
 1201 tgtgaagccc gagtgcgatt cagaggcgga ggacagtggc accgaggatg ccgtagacgc  
 1261 ggagctggag cacatggaac tagactttga gtgcggtggg aaccgaagcg gtggaagcga  
 1321 ttttctatc aatgagggcg tctttgaaca ggatcttct accgatgtgc agtgtgcctt  
 1381 tcatgtgcaa ccgccgactt tggccactc gtatttaaat attcattatg tgtgtgagac  
 50 1441 gggctcgcga atcattttt tcaccatcca tacccttga aaggttccag tttcgaaca  
 1501 attggaagcc catacacagg tgaactcct gagaggagtg tggccagcat taatggctat  
 1561 agctttggcg cagtgcagg gtcagcttc ggtgccacc attatcgggc agttattca  
 1621 aagcactcgc cagctagcgg atatcgataa gatcgaaccg ttgaagatct cgaagatggc  
 1681 aaatctcacc aggacctgc acgactttgt ccaggagctc cagtactgg atgttactga  
 55 1741 tatggagttt ggctgtctgc gtctgactt gctcttcaat ccaacgctct tgcagcagcg

1801 caaggagcgg tcgttgcgag gctacgtccg cagagtccaa ctctacgctc tgtcaagttt  
 1861 gagaaggcag ggtggcatcg gcggcggcga ggagcgcttt aatgttctgg tggctcgctt  
 1921 tcttcgctc agcagcctgg acgagagggc catggaggag ctgttctcg ccaacttgg  
 1981 ggggcagatg cagatggatg ctcttattcc gttcactatg atgaccagca acaccagtgg  
 5 2041 actgtaggcg gaattgagaa gaacagggcg caagcagatt cgctagactg cccaaaagca  
 2101 agactgaaga tggaccaagt gcgggcaata catgtagcaa ctaggcaaat cccattaatt  
 2161 atatatttaa tatatacaat atatagttaa ggatacaata ttctaacata aaacatggg  
 2221 tttattgttg ttcacagata aatggaatc gatttcccaa taaaagcgaa tatgtttta  
 2281 aacagaat

10

**27. SEQ ID NO: 27 Accession No. NM\_057433 Drosophila melanogaster  
 ultraspiracle CG4380-PA (usp)**

MDNCDQDASFRLSHIKEEVKPDISQLNDSNNSSFSPKAESPVPF  
 15 MQAMSMVHVLPGSNSASSNNNSAGDAQMAQAPNSAGGSAAAAVQQQYPPNHPLSGSKH  
 LCSICGDRASGKHVYVSCEGCKGFFKRTVRKDLTYACRENRNCCIHDRQRNRCQYCR  
 YQKCLTCGMKREAVQEERQRGARNAAGRLSASGGGSSGPGSVGGSSSQGGGGGGGGVSG  
 GMGSGNGSDDFMTNSVSRDFSIERIIEAEQRAETQCGDRALTFLRVGPYSTVQPDYKG  
 20 AVSALCQVVNKQLFQMVEYARMMPHFAQVPLDDQVILLKAAWIELLIANVAWCSIVSL  
 DDGGAGGGGGGLGHDGSFERRSPGLQPQQLFLNQSFYSYHRNSAIKAGVSAIFDRILSE  
 LSVKMKRLNLDRELSCLKAILYNPDIRGIKSRAEIECMCREKVYACLDEHCRLEHPG  
 DDGRFAQLLLRLPALRSISLKCQDHLFLFRITSDRPLEELFLEQLEAPPPPGLAMKLE

25

**28. SEQ ID NO: 28 Accession No. NM\_057433 Drosophila melanogaster  
 ultraspiracle CG4380-PA (usp)**

1 aaaaatgtcg acgcgaaaaa aggtatttat tcattagtea gaaagtctgg cattctttgt  
 61 ttgttgtaa aaagecgaat tgttgagg cgagcgaata aagtgcgtg ctccatcggc  
 121 tcaagattat gtaaatgcag caacgacccc accaacaacg aaactgcaac ctgctccact  
 30 181 tggcccaacg gaccaatagc ggacggacgg acacgggtggc gttggcaaag tgaacccca  
 241 acagagaggc gaaagcgagc caagacacac cacatacaca cgaagagaac gagcaagaag  
 301 aaaccggtag gcggaggagg cgctgcccc agttctcca atataccag caccacatca  
 361 caagcccagg atggacaact gcgaccagga cgccagcttt cggtgagcc acatcaagga  
 421 ggaggtcaag cggacatct cgcagtgaa cgacagcaac aacagcagct ttgcccga  
 35 481 ggccgagagt cccgtgccct tcattcaggc catgtccatg gtccacgtgc tgcccggtc  
 541 caactccgc agtccaaca acaacagcg cgagatgcc caaatggcg aggcgcccga  
 601 ttcggctgga ggctctgcc cgcgtcagc ccagcagcag tatccgcta accatccgt  
 661 gagcggcagc aagcacctct gctctattg cggggatcg gccagtggca agcactacgg  
 721 cgtgtacagc tgtgagggt gcaagggct cttaaacgc acagtgcga aggatctac  
 40 781 atacgttgc agggagaacc gcaactgcat catagacaag cggcagagga accgtgcca  
 841 gtactgccg taccagaagt gcctaactg cggcatgaag cgcgaagcgg tccaggagga  
 901 gcgtcaacgc ggcgcccga atgcggcggg taggtcagc gccagcggag gcggcagtag  
 961 cgtccaggt tcgtaggcg gatccagctc tcaaggcgga ggaggaggag gcggcggtt  
 1021 tggcggaatg ggcagcggca acggttctga tgacttcatg accaatagcg tgtccaggga  
 45 1081 ttctcgatc gagcgcatca tagaggcca gcagcgagcg gagacccaat gcggcgatc  
 1141 tgactgacg ttctgcgcg ttgtcccta ttccacagc cagccggact acaagggtg  
 1201 cgtgtcgcc ctgtccaag tggtaacaa acagctctc cagatggctg aatacgcgc  
 1261 catgatccg cacttgccc aggtgccgt ggacgaccag gtgattctg tgaagccgc  
 1321 ttgatcgag ctgtcattg cgaacgtggc ctggtcagc atcgttgcg tggatgacg  
 50 1381 cggtgccggc ggcggggcg gtggactagg ccacgatggc tccttgagc gacgatcacc  
 1441 gggcctcag cccagcagc tttctcga ccagagctc tctaccatc gcaacagtgc  
 1501 gatcaaagc ggtgtgtcag ccatctcga ccgcatattg tcggagctga gtgtaaagat  
 1561 gaagcggctg aatctcgacc gacgcgagct gtctgctg aaggccatca tactgtacaa  
 1621 cccggacata cgcgggatca agagccggcg ggagatcgag atgtcccgcg agaaggtga  
 55 1681 cgcttgctg gacgagcact gccgcctgga acatccggcg gacgatggac gcttgcgca

1741 actgctgctg cgtctgccg ctttgcgac gatcagcctg aagtgccagg atcacctgtt  
 1801 cctcttcgc attaccagcg accggccgct ggaggagctc ttctcgagc agctggagge  
 1861 gccgccgcca cccggcctgg cgatgaaact ggagtagggt cccgactcta aagtctccc  
 1921 cgttctccat ccgaaaaatg ttctatttg attgcgttg ttgcatttc tctctctat  
 5 1981 cccttatacc ctacaaaagc cccctaatac tacgcaaaat gtgtatgtaa ttgtttattt  
 2041 ttttttatt acctaatac attattatta ttgatataa aaatgtttc ctaagatga  
 2101 agattagcct cctcgacgtt tatgtcccag taaacgaaaa acaaacaaaa tccaaaactt  
 2161 gaaaagaaca caaacacga acgagaaaaat gcacacaagc aaagtaaaag taaaagttaa  
 2221 actaaagcta aacgagtaaa gatattaaaa taacggtaa aattaatgca tagttatgat  
 10 2281 ctacagacgt atgtaacat acaaatcag cataaatata tatgtcagca ggcgcatac  
 2341 tgcggtgctg gccccgtct aaatcaattg taattacttt ttaacataaa ttaccctaaa  
 2401 acgttatcaa ttagatgca gatacaaaaa tcaccgacga aaaccaacaa aatatactca  
 2461 tgtataaaaa atataaactg cataacaa

15 **29. SEQ ID NO: 29 Accession No. NM\_168757 Drosophila melanogaster**  
**Ecdysone-induced protein 75B CG8127-PD**

MGEELPILKGILKGNVNYHNAPVRFGRVPKREKARILAAMQQST  
 QNRGQQRALATELDDQPRLLAAVLRAHLETCEFTKEKVSAMRQRARDCPSYSMPULLA  
 20 CPLNPAPELQSEQFSQRFHVIRGVDFAGMIPGFQLLTQDDKFTLLKAGLFDALFV  
 RLICMFDSSINSIICLNGQVMRRDAIQNGANARFLVDSTFNFAERMNSMNLTD AEIGL  
 FCAIVLITPDRPGLRNLELIEKMYSLKGLQYTVAQNRPDQPEFLAKLLETMPDLRT  
 LSTLHTEKL VVFRTEHKELLRQQMWSMEDGNNSDGQQNKSPSGSWADAMDVEAAKSPL  
 GSVSSTESADLDYGSPPSSQPQGVSLPSPQQQPSALASSAPLLAATLSGGCPLRNRA  
 25 NSGSSGDSGAAEMDIVGSHAHLTQNGLTITPIVRHQQQQQQQQIGILNNAHSRNLNG  
 GHAMCQQQQQHPQLHHHLTAGAARYRKLDSPDTSIESGNEKNECKAVSSGGSSSCSS  
 PRSSVDDALDCSDAAANHNQVVQHPQLSVSVSPVRSPQPSTSSHLKRQIVEDMPVLK  
 RVLQAPPLYDTNSLMDEAYKPHKKFRALRHREFETAADASSSTSGSNSLSAGSPRQS  
 PVPNSVATPPPSAASAAAGNPAQSQLHMHLTRSSPKASMASHSVLAKSLMAEPRMTP  
 30 EQMKRSDIIQNYLKRENSTAASSTTNGVGNRSPSSSSTPPPSAVQNQQRWGSSSVITT  
 TCQQRQQSVSPHSNGSSSSSSSSSSSSSSSTSSNCSSSSASSCQYFQSPHSTSNGT  
 SAPASSSSGSNSATPLLELQVDIADSAQPLNLSKKSPTPPPSKLHALVAAANAVQRYP

TLSADVTVTASNGGPPSAASPAPSSPPASVGSPNPGLSAAVHKVMLEA

35 **30. SEQ ID NO: 30 Accession No. NM\_168757 Drosophila melanogaster**  
**Ecdysone-induced protein 75B CG8127-PD**

1 agtcaccgtc gcagtcgcag cagttgaggt tcgctctcct cgatttcggg caaatccgat  
 40 61 accatatagc acagcgtacc gcactctggg tatattcgta acgcgcttg gcttttacag  
 121 ttatgcgctg tcgagacctt gtcgagttt gtcattgtag ccagcgatcc gcgggatccg  
 181 aaataagcca agaatacaaa cgcgagtgcg gcagttgcca gcagtaacta caccaatatt  
 241 tatattaatt aaaataaatt aaatgaaaca acatgctgat taatgccaat gaatgttaa  
 301 tgcaattgtt aatgtgaaga aaagtcgacc aagtctccc aaaacaacac ttattcaaca  
 45 361 tccactacac actgccttt ctggattacg cgcccaaaaa aaaacaaaaa ttaaaaatta  
 421 aaccaaacca acaactaatt tatttgctaa atatccaaa aattcaatca atgtgaaaag  
 481 caagcaaaca aagttcctet cacaacaaaa cagcagttaa taaaatate taaccgagat  
 541 aaagtgcaca gaagataaca agtttctcaa gcaaacatcc atatgtacct gattaccaac  
 601 caaaaagctg tgtgtgtgcc aaaaaccgaa gaggaattat ccaaaaatat ttaatgagca  
 50 661 agtcaactg agtgggtgat gtgccccca agggaaaagt gaccaagtca agatattttg  
 721 tcaaatcgaa cacagaaaac aaaaaatgg gcgaagaact cccgatattg aagggcatac  
 781 ttaaaggcaa cgtcaactat cacaatgcgc ctgtgcgtt tggacgcgtg ccgaagcgcg  
 841 aaaaggcgcg tatcctggcg gccatgcaac agagcaccca gaatcgcggc cagcagcgag  
 901 ccctcgccac cgagctggat gaccagccac gcctctcgc cgccgtgctg cgcgccacc  
 55 961 tcgagacctg tgagttcacc aaggagaagg tctcggcgat gcggcagcgg gcgcgggatt

1021 gcccctccta ctcctatgcc acacttctgg cctgtccgct gaaccccgcc cctgaactgc  
 1081 aatcggagca ggagtctcgc cagcgtttcg cccacgtaat tcgcggcgtg atcgactttg  
 1141 ccggcatgat tcccggcttc cagctgtca cccaggacga taagttcacg ctctgaagg  
 1201 cgggactctt cgacgccctg tttgtgcgc tgatctgcat gtttgactcg tcgataaact  
 5 1261 caatcatctg tctaaatggc caggtgatgc gacgggatgc gatccagaac ggagccaatg  
 1321 cccgcttcct ggtggactcc accttcaatt tcgcggagcg catgaactcg atgaacctga  
 1381 cagatgccga gataggcctg ttctgcgcca tcgttctgat tacgccggat cgccccggtt  
 1441 tgcgcaacct ggagctgacg gagaagatgt actcgcgact caagggctgc ctgcagtaca  
 1501 ttgtcgcca gaataggccc gatcagccc agttcctggc caagttgctg gagacgatgc  
 10 1561 ccgatctgcg caccctgagc accctgcaca ccgagaaact ggtagttttc cgcaccgagc  
 1621 acaaggagct gctgcgccag cagatgtggt ccatggagga cggcaacaac agcgatggcc  
 1681 agcagaacaa gtcgcctcgc ggcagctggg cgcatgcat ggacgtggag gcggccaaga  
 1741 gtccgcttgg ctcggtatcg agcactgagt ccgccgacct ggactacggc agtccgagca  
 1801 gttcgcagcc acagggcgtg tctctgccct cggcgctca gcaacagccc tcggctctgg  
 15 1861 ccagctcggc tctctgctg gcggccaccc tctccggagg atgtcccctg cgcaaccggg  
 1921 ccaattccgg ctccagcggg gactccggag cagctgagat ggatatcgtt ggctcgacg  
 1981 cacatctcac ccagaacggg ctgacaatca cgccgattgt gcgacaccag cagcagcaac  
 2041 aacagcagca gcagatcgga atactcaata atgcgcattc ccgcaacttg aatgggggac  
 2101 acgcgatgtg ccagcaacag cagcagcacc cacaactgca ccaccacttg acagccggag  
 20 2161 ctgcccgtca cagaaagcta gattcgcca cggattcggg cattgagtcg ggcaacgaga  
 2221 agaacgagtg caaggcgtg agttcggggg gaagttctc gtgtccagt ccgcgttcca  
 2281 gtgtggatga tgcgctggac tgcagcgatg ccgccgcaa tcacaatcag gtggtgcagc  
 2341 atccgcagct gagtgtggtg tccgtgtcac cagttcgtc gcccagccc tccaccagca  
 2401 gccatctgaa gcgacagatt gtggaggata tgcccgtgct gaagcgcgtg ctgcaggctc  
 25 2461 cccctctgta cgataccaac tcgctgatgg acgaggccta caagccgcac aagaaattcc  
 2521 gggccctgcg gcacgcgag ttcgagaccg ccgaggcgga tgccagcagt tccattccg  
 2581 gctcgaacag cctgagtgcg ggcagtcgc gacagagtc agtcccgaa agtgtggcca  
 2641 cgccccgccc atcgggcgcc agcgccgccc caggtaatcc cgcccagagc cagctgcaca  
 2701 tgcacctgac ccgagcagc cccaaggcct cgatggccag ctgcactcg gtgtggcca  
 30 2761 agtctctcat ggccgagccg cgcatgacgc ccgagcagat gaagcgcagc gatattatcc  
 2821 aaaactactt gaagcgcgag aacagcacag cagccagcag caccaccaat ggcgtgggca  
 2881 accgcagtc cagcagcagc tccacaccgc cgccatcggc ggtccagaat cagcagcgtt  
 2941 ggggcagcag ctcggtgac accaccacct gccagcagcg ccagcagtc gtgtcgccgc  
 3001 acagcaacgg ttccagctcc agttcgagct ctactccag ctccagttcg tcactctct  
 35 3061 ccacatctc caactgcagc tccagctcgg ccagcagctg ccagtattc cagtcgccgc  
 3121 actccaccag caacggcacc agtgaccgg cgagctccag ttggggatcg aacagcgcca  
 3181 cgccctgtg ggaactgcag gtggacattg ctgactcggc gcagcctctc aattgtcca  
 3241 agaaatcgcc cagccgccc cccagcaagc tgcacgctct ggtggccgccc gccaatgccg  
 3301 ttcaaaggta tcccacattg tccgccagc tcacagtac agcctccat ggcggtctc  
 40 3361 cgtcggcgcc ggcgagtcg gcgcccagca gcagtcgccc ggcgagtggt ggctcccca  
 3421 atccgggcct gagcgccgccc gtgcacaagg taatgtgga ggcgtaagag cgggaggagg  
 3481 tagtggttt tacgcggaga agtgggagag acagagactg ggagtggcag ttacgcgaag  
 3541 caggaagcag gatcacttgg agcgggcgga gttgaattaa attattttac catttaattg  
 3601 agacgtgtac aaagttgaa agcaaaacca acatgcatgc aatttaaac taatatttaa  
 45 3661 agcaacaaca acaaaaaca ctacaagta ttaatttaa aaacaacaa acaaacaaac  
 3721 acaaaaaaac ccaagcttga atggtattac

### 31. SEQ ID NO: 31 Accession No. NM\_168892 *Drosophila melanogaster*

#### Ecdysone-induced protein 78C CG18023-PBEip78C)

50

MHPSHLQQQQQQHLLQQQQQQQHQPQLQQHHQLQQQPHVSGVRV  
 KTPSTPQTPQMCSIASSPSELGGCNSANNNNNNNNNSSSGNASGGSGVSVGVVVVGGH  
 QQLVGGSMVGMAGMGTDHQAQVGMCHDGLAGTANELTVYDVIMCVSQAHLNCSYTEEL  
 TRELMRRPVTPQNGIASTVAESLEFQKIWLWQQFSARVTPGVQRIVEFAKRVPGFCD  
 55 FTQDDQLILIKLGFEEVWLTHVARLINEATLTLDDGAYLTRQQLEILYDSDFVNALLN  
 FANTLNAYGLSDTEIGLFSAMVLLASDRAGLSEPKVIGRARELVAEALRVQILRSRAG

SPQALQLMPALEAKIPELRSLGAKHFSHLDWLRMNWTKLRLPPLFAEIFDIPKADDEL

**32. SEQ ID NO: 32 Accession No. NM\_168892 Drosophila melanogaster  
Ecdysone-induced protein 78C CG18023-PBEip78C)**

5

10

15

20

25

30

35

40

45

50

55

1 aagcattaac gaaagaactg cgcacaaagt agggaggcaa taattacata tgtacatggc  
 61 tgggaaaggc cttactaaa cttagcaaac taataaatag aaaaaaggaa atattggcca  
 121 aatattatag tattgggaat attagggtac ttgatatcaa aaattaatgt ctattttata  
 181 cacttattct tagacttaac gtaacttat cgtacttatt atgattgggt ttcaagatt  
 241 accagaactt gatagattgg tctagctttt gaaatcggat agcattttct ttaaaggact  
 301 ttgccatag ctaaaccta acttctttt tcaattcagc cacagctgac aaaagcgaag  
 361 aaaattgaa agaccgtgaa tcttttgaa acgccctctc cggattcctc attaagtga  
 421 aaagatataa catcgcagag atttccata aaaatgctga tcaggcgccc tcgcagggtg  
 481 ccaacgtcga ttccgccag caggacgatg atgaagatga tggatgccc tctaccgat  
 541 tcatccgag caacatggat gtatacaaaa tagagctgga ggaacaggca caaatccgt  
 601 ccaaactgct ggtcgaaacc tgtgtgaagc actcgtcttc ggagcagcag cagctccaag  
 661 ttaagcagga ggacctcatc aaggatttca ctccgggacga ggaggaacag ccaagcgaag  
 721 aggaggcgga ggaagaggac aacgaagagg acgaggaaga agaaggcgaa gaagaagagg  
 781 aggacgagga cgaggaagcc ctgctgccgg tagtcaattt taatgcaaat tcagacttta  
 841 atttgcattt cttgacaca ccggaggact cgtccacca aggggcctac agtgaggcca  
 901 atagcttga atccgagcag gaagaggaga agcaaacaca gcagcatcag cagcagaagc  
 961 agcatcaccg ggatttggag gattgcctaa gtgccattga agctgatcca ttgcagttgt  
 1021 tgcattgca cgaattctat agaacatcag cctagcaga gagtgttga gccagtctaa  
 1081 gccacagca gcagcagcaa cggcagcaca cccaccagca acaacagcaa cagcagcagc  
 1141 agcagcaaca cctggacag cagcaacatc agctcaactg cagctgagc aatggtggag  
 1201 gtgctttga caccatcagc agtgtgcatc agttcggtcc ggccagcaac cacaacacca  
 1261 gcagcagctc cccctcctcc agcgccgccc actcttcgcc ggacagcggc tgcctcgtcg  
 1321 cctcctctc cggatcttcg cgtacctgag gatcctctc tgcactctc tctcgtcag  
 1381 cggtcagcag caccatcagc agcgcccgca gcagcaaca cagcgtcgtc aaccccgag  
 1441 caacatctc atctgttgc catctgaaca aagagcaaca gcagcagcca ctgccgacga  
 1501 cacagctgca acagcagcag cagcaccagc agcagttgca acaccgcag cagcagcaat  
 1561 cttttggcct agcagacagc agcagcagca acggcagcag caacaacaac aacggtgtct  
 1621 cctcgaaatc atttgtgccc tgcaaatgtc gtggcgacaa ggcatcgga taccactatg  
 1681 gtgtaacctc ctgcgagggt tgcaagggt tcttctgctc cagtatccag aagcaaatcg  
 1741 aatatcgtg ttgcgggac ggcaagtgcc tggatcag actgaaccgc aatcgtgcc  
 1801 agtactgccg ctcaagaaa tgccttccg ctggcatgag ccgcgattcc gtacgttatg  
 1861 gtcggttcc caagcgttcc cgtgagctga acggagcggc cgcctcctcc gccgccgtg  
 1921 gagctcctgc ctccctcaat gtggatgact ctaccagcag cactatgcac ccagtcacc  
 1981 tacagcagca gcagcaacag catctactac agcagcaaca gcagcagcaa catcagccac  
 2041 agctgcagca acaccacaa ctgcaacagc agccgcatgt aagcggcgta cgtgtgaaga  
 2101 ccccgagtac tccacaaacg ccacaaatgt gttcgatcgc ctctcgcga tcggagctgg  
 2161 gcggttgcaa tagtgcaat aacaataaca ataataaca caacagtagc agcggtaatg  
 2221 ccagcgttgg cagcggtg agcgtcggcg ttgtgtgtg gggcgacac cagcaactgg  
 2281 tgggaggcag catgttggga atggcgga tgggcacgga tgcccaccag gtgggcatgt  
 2341 gtcacgacgg cttggcgga acggcaaacg agctgaccgt ctacgatgc atcatgtgcg  
 2401 tgtcgcaggc gcaccgctc aactgctct acacggagga actgaccaga gagtcatgc  
 2461 gtcgtccgt gacggtgcca caaatggga ttgccagcag agtggccgag agtctggagt  
 2521 tccagaagat ctggctgtgg caacagttct cggccagggt gacgcctggc gttcagcga  
 2581 ttgtggagt tgcaaacgc gtacctggt tctgtgatt caccgaagat gaccagctta  
 2641 tactaataaa gctgggctt ttcaggtct ggttgacca tgtggccgg ttgatcaatg  
 2701 aggcgacatt gacactggac gatgtgctt acctgacgc ccagcagctt gagatactct  
 2761 acgattctga cttgtcaac gccttctga acttgccaa cacgtgaac gcctacgggc  
 2821 tgagtgcac cgaaatcgga ctctctcgg ccatggtgt gcttgcctc gatcagctg  
 2881 gactcagcga gccaagggt atcggcagg ccagggaact ggtggccgag gcgtgcgcg  
 2941 tacagatcct gcgttcgag gcagatccc cacaggcgt gcagctgat ccggcgctgg  
 3001 aagccaagat acccgagctg agatccttgg gggccaagca cttctcacac ctgactggc

3061 tacggatgaa ctggaccaag ctgcgcctgc cggccctctt cggcgagatc ttcgacatcc  
 3121 cgaaggctga cgatgagctg taggatgtgg agccaacccc gcgattccag ggccgtgcaa  
 3181 agcaaaccgc aacaagaaca gaattattta ccaattgtag gcttaagcaa cgtagctata  
 3241 gatcgaaatg ggagggccgc agatcagata cacgtctact cagcattacc ggagagatag  
 5 3301 tccactaagc ctatatgcat actactatac tagcagtgtt a

**33. SEQ ID NO: 33 Accession No. NM\_165465 Drosophila melanogaster  
 Ecdysone receptor CG1765-PB (EcR)**

10 MKRRWSNNGGFMRLPEESSSEVTSSSNGLVLP SGVNMSPSSLDS  
 HDYCDQDLWLCGNESGSFSGSNGHGLSQQQSVITLAMHGCSSTLPAQTTIIPINGNA  
 NGNGGSTNGQYVPGATNLGALANGMLNNGGFNGMQQQIQNGHGLINSTTPSTPTTPLHL  
 QQNLGGAGGGGGIGGMGILHHANGTPNGLIGVVGGGGGVGLGVGGGGVGGGLGMQHTPRS  
 DSVNSISSGRDDLSPSSSLNGYSANESCDAKKSKKGPAPRVQEELCLVCGDRASGYHY  
 15 NALTCEGCKGFFRRSVTKSAVYCKKFRACEMDMYMRRKCQECRLKKCLAVGMRPECV  
 VPENQCAMKRREKKAQKEKDKMTTSPSSQHGGNGSLASGGGQDFVKKEILDLMTCPEP  
 QHATIPLLPDEILAKCQARNIPSLTYNQLAVIYKLIWYQDGYEQPSEEDLRRIMSQPD  
 ENESQTDVSRHITEITILTVQLIVEFAKGLPAFTKIPQEDQITLLKACSSEVMMLRM  
 ARRYDHSSDSIFFANNRSYTRDSYKMAGMADNIEDLLHFCRQMFSMKVDNVEYALLTA  
 20 IVIFSDRPGLEKAQLVEAIQSYIDTLRIYILNRHCGDSMSLVFYAKLLSILTELRTL  
 GNQNAEMCFSLKLNKRLPKFLEEIWDVHAIPPSVQSHLQITQEENERLERAERMRAS  
 VGGAITAGIDCDSASTSAAAAAAQHQPQPQPQPSSLTQND SQHQTQPQLQPQLPPQ  
 LQGQLQPQLQPQLQTQLQPQIQPQPQLLPVSAPVPASVTAPGSLSAVSTSSEYMGGSA  
 AIGPITPATTSSITA AVTASSTTSAVPMGNGVGVGVGVGGNVSMYANAQTAMALMGVA  
 25 LSHHQEQLIGGVAVKSEHSTTA

**34. SEQ ID NO: 34 Accession No. NM\_165465 Drosophila melanogaster  
 Ecdysone receptor CG1765-PB (EcR)**

30 1 tagtattttt ttggactttg ttgttaacgg ttgttcgctc gcacgtacga agcccgatcg  
 61 cgttcgtcaa aaaacaagat acaaaataca gcacacacaa ttgaaaacga caacctaaca  
 121 gtacggtttc ccaaagcacc ttacatttca aaaccgaaaa ccccaaaat ttgttaacca  
 181 aataatgttt aatcacata tacacctaca tatatttatg aaaaattgtt agacaaatcc  
 35 241 caaataatac cagttccccc aacaaccgca acaaacacaa gtgcaattca tcggcaaaaa  
 301 ttaataataa gtgcaaatgc attgtagctg aaactcaaac aatagtaaaa atacatacat  
 361 aagtggtgaa gaagcaaaag gaaatagttc ttaaaataac gcaaatcgag agcatatatt  
 421 catatttgta cagatattat atggcggctg catagtgcga actgcggctg agggataaca  
 481 gcggtatcga aatgtaaata ggaaacaacg aagccagaac tcgaaatcaa acatcagcaa  
 40 541 cgtgacacac agacataaga cgcccgctca gtcgtggtct gtggaacgct agctccgctt  
 601 tgccaggagc cggagacttt ttccgcatcc acaatattac atatgtacat atacgaaga  
 661 tagtgccgca gtgagtgagg gatttgtgcc gtggatcccg atccccctac atatataata  
 721 aggtagtga aagattttac tcaacattcc aaatagtgt ttgtcaactg gaataccttt  
 781 tgttcaata cgcagtgggc ccatggatac ttgtggatta gtagcagaac tggcgcacta  
 45 841 tatcgacgca tatgctctga ttgtttcccg cactaaatga gcagggattc gggcgaaaat  
 901 gtatttigaa cgcaacaag tgcgcaaaaa atactagtc caccacgaaa ctgcacaaaa  
 961 caccgccaga agcgagcaga acctcgggcc gcacgaccga gcttcgtaaa gcaacagagg  
 1021 atcttaccag gagatagctc ttctccacat agaccaactg ccagggacaa gctcctgtc  
 1081 ccagccgac gctaagtga cggaaaacgg ccacaaaacg gcgactatcg gctgccagag  
 50 1141 gatgaagcgg cgctggctga acaacggcgg ctcatgcgc ctaccggagg agtcgtctc  
 1201 ggaggtcacg tctctctga acgggctcgt cctgccctcg ggggtgaaca tgcgccctc  
 1261 gtcgtggac tcgcacgact attgcgatca ggaccttgg ctctgcggca acgagtccgg  
 1321 ttctgttggc ggctccaacg gccatggcct aagtcagcag cagcagagcg tcatcacgt  
 1381 ggcatgcac ggggtgtcca gcactctgcc cgcgcagaca accatcatc cgatcaacgg  
 55 1441 caacgcgaat gggaatggag gctccaccaa tggccaatat gtgccgggtg ccactaatct

1501 gggagcgttg gccaacggga tgctcaatgg gggcttcaat ggaatgcagc aacagattca  
 1561 gaatggccac ggcctcatca actccacaac gccctcaacg ccgaccaccc cgctccacct  
 1621 tcagcagaac ctggggggcg cggggcggcg cggatcggg ggaatggga ttctcacca  
 1681 cgcgaatggc accccaaatg gccttatcg agttgtggga ggcggcggcg gagtaggtct  
 5 1741 tggagtaggc ggaggcggag tgggaggcct gggaatgcag cacacacccc gaagcgattc  
 1801 ggtgaattct atatcttcag gtcgcgatga tctctgcct tcgagcagct tgaacggata  
 1861 ctcggcgaac gaaagctgcg atgcgaagaa gagcaagaag ggacctgcgc cacgggtgca  
 1921 agaggagctg tgcttggtt gcggcgacag ggcctccggc taccactaca acgccctcac  
 1981 ctgtgagggc tgcaaggggt tcttcgacg cagcgttacg aagagcgccg tctactgtg  
 10 2041 caagttcggg cgcgcctgcg aatggacat gtacatgagg cgaaagtgc aggagtgcg  
 2101 cctgaaaaag tgcttgccg tgggtatgcg gccggaatgc gtcgtcccgg agaaccaatg  
 2161 tgcgatgaag cggcgcgaaa agaaggccca gaaggagaag gacaaaatga ccacttcgcc  
 2221 gagctctcag catggcggca atggcagctt ggcctctggt ggcggccaag actttgttaa  
 2281 gaaggagatt ctgacctta tgacatgcga gccgcccag catgccacta tccgctact  
 15 2341 acctgatgaa atattggcca agtgtcaagc gcgcaatata ccttcctaa cgtacaatca  
 2401 gttggccgtt atatacaagt taatttggtt ccaggatggc tatgagcagc catctgaaga  
 2461 ggatctcagg cgtataatga gtcaaccga tgagaacgag agccaaacgg acgtcagctt  
 2521 tcggcatata accgagataa ccatactcac ggtccagttg attgttgagt ttgctaaagg  
 2581 tctaccagcg ttacaaaga taccacagga ggaccagatc acgttactaa aggcctgtc  
 20 2641 gtcggaggtg atgatgtgc gtatggcagc acgtatgac cacagctcgg actcaatatt  
 2701 ctctcgcaat aatagatcat atacgcggga ttctacaaa atggccgga ttgctgataa  
 2761 cattgaagac ctgctgcat tctccgcca aatgtctcg atgaagggtg acaacgtcga  
 2821 atacgcgctt ctactgcca ttgtatctt ctggaccgg cgggcctgg agaaggccca  
 2881 actagtcgaa gcgatccaga gctactacat cgacacgcta cgcattata tactcaaccg  
 25 2941 ccactgcggc gactcaatga gcctcgtct ctacgcaaag ctgctctcga tctcaccga  
 3001 gctgcgtacg ctgggcaacc agaacgccga gatgtgttc tactaaagc tcaaaaaccg  
 3061 caaactgcc aagtctctg aggagatctg ggacgttcat gccatccgc catcggcca  
 3121 gtcgcacctt cagattacce aggaggagaa cgagcgtctc gagcgggctg agcgtatgcg  
 3181 ggcatcgggt gggggcgcca ttaccgccgg cattgattgc gactctgctt ccacttcggc  
 30 3241 ggcggcagcc gcggcccagc atcagcctca gcctcagccc cagccccaac cctcctccct  
 3301 gaccagaac gattcccagc accagacaca gccgcageta caacctcagc taccacctca  
 3361 gctgcaaggt caactgcaac ccagctcca accacagctt cagacgcaac tccagccaca  
 3421 gattcaacca cagccacagc tcttcccgt ctccgctccc gtgcccgcct ccgtaaccgc  
 3481 acctggttcc ttgtccgcg tcagtaagag cagcgaatac atggggcgga gtgcggccat  
 35 3541 aggaccatc acgccggcaa ccaccagcag taccacggct gccgttaccg ctagctccac  
 3601 cacatcagcg gtaccgatgg gcaacggagt tggagtcggt gttggggtgg gcggcaacgt  
 3661 cagcatgtat gcgaacgcc agacggcgat ggccttgatg ggtgtagccc tgcattcgca  
 3721 ccaagagcag ctatcgggg gagtggcggg taagtcggag cactcgacga ctgcatagca  
 3781 ggcgcagagt cagctccacc aacatcacca ccacaacatc gacgtcctgc tggagtagaa  
 40 3841 agcgcagctg aaccacaca gacatagggg aaatggggaa gttctctcca gagagttcga  
 3901 gccgaactaa atagtaaaaa gtgaataatt aatggacaag cgtaaatgc agttattag  
 3961 tcttaagcct gcaaatatta cctattatc atacaaatta acatataata cagcctatta  
 4021 acaattacgc taaagcttaa ttgaaaaagc ttcaacaaca attggacaaa cgcgttgagg  
 4081 aaccgggaga aaatttaaga aaaaaaaaaa cattgaaaat tatgaaattt agtatacat  
 45 4141 tttttgggt ggatgtatgt cgcacagac tcacgatcaa ttctcgaatt ttgttaacta  
 4201 aattgacct ccaaactgca tgcgaacag atcagaaaaag agaacagaca gtaggcgctg  
 4261 aacagaggga agagagaaga gaataaagat tgtttatatt taaaaatat ataaataat  
 4321 aattactaac tctaaacgta atgaaagcaa ctgtataata tctaactata actataaatt  
 4381 cgtactgtag ggaagtgaga aaatctgtta aatgaaacaa aaataatgat aataacatta  
 50 4441 tcatccacca taattaaat catttaaagt aattaaaaac aaaacacttt taaaacacgc  
 4501 aaaacttgga ctgattttat aaatattttt taatcataaa gaaaggcaac ctgaaaaaaa  
 4561 tattacaaa acaataaca acatatttta ttatgacacc ctatattgtt ttcaaacga  
 4621 gaatttaaatt tcttagattc ttataattc atccaaaaat attagccagc aaaaacctt  
 4681 attattggca ttgttttag acatgtttc aaaaaaaact ttgatattga aactaaacaa  
 55 4741 aggataatga aatgaaagt attggagtct tactcaaaaa ccaaaggca tcaaaaggta  
 4801 ttaaattaaa aatataatct aatttcagat tcaagaaaca cttttgggtg gaaaatagtt  
 4861 ttcaatcact ttgataaaaa ccacacaaat taataatac atgcatacac caaagactt  
 4921 caatatatat ttttaaaatt tacattgata attcgaaatt tgaataagaa tcacatccat

4981 ctaatttggc taaatcaaaa ttttatgaa agccacacaa aaaacgtgca aatttgatta  
 5041 ctttggcaat ttttatgta tacaaaatit atgcaattga tttcaaaat aatttttatt  
 5101 agattgtatt agtttcattt tgctttggga tgtacatttt aaataaattt tactttaaat  
 5161 tgttggcctt attttaactt aaatcaaatt tattctaatt ttagtaaaaa aaaatgtgtt  
 5 5221 taaaattgaa aataagaaca ctgtaaaata ttaataaaaa attaaagttt aaagtgttc  
 5281 ttttattatg taaaaagaag acaaaaaata tcttacgtag ctttctactt gaattgtgca  
 5341 atttttactt ttactacta atcctaattt aaatataatt tacacacacg cctacacatc  
 5401 cageccacata tttttaattt taagtcaacc taatttataa atatgaattt gtataatgac  
 5461 gaactaaaat tagcatgaca tcatggacat acttggaat aactctatca aacgagctaa  
 10 5521 atgcattgaa gaagaaaatt cttgttaaatt atagtctgca cttcgacaaa cgaaaatcag  
 5581 tgaatt

**35. SEQ ID NO: 35 Accession No. NM\_165364 Drosophila melanogaster**  
**Hormone receptor-like in 39 CG8676-PD Hr39)**

15 MPNMSSIKAEQQSGPLGGSSGYQVPVNMCTTTTVANTTTTLGSSA  
 GGATGSRHNVSVTNIKCELDELPSNGNMVPIANYVHGSLRIPLSGHSNHRES DSEE  
 ELASIENLKVRRRTAADKNGPRPMSWEGELSDTEVNGGEELMEMEPTIKSEVVPVAVAP  
 PQPVCALQPIKTELENIAGEMQIQEKYCYPQSNTQHHAATKLKVAPTQSDPINLKFEPP  
 20 LGDNSPLLAARSKSSSGGHLPLPTNPSPDSAIHSVYTHSSPSQSPLTSRHAPYTPSLS  
 RNNSDASHSSCYSYSSEFSPTHSPIQARHAPPAGTLYGNHHGIYRQMKVEASSTVPSS  
 GQEAQNLSMDSASSNLDTVGLGSSHPASPAGISRQQLINSPCPICGDKISGFHYGIFS  
 CESCKGFFKRTVQNRKNYVCVRGGPCQVSISTRKKCPACRFKCLQKGMKLEAIREDR  
 TRGGRSTYQCSYTLPN SMLSPDQAAAAAAAAAAVASQQQPHQRLHQLNGFGGVPI  
 25 PCSTSLPASPSLAGTSVKSEEMAETGKQSLRTGSVPPLLQEIMDVEHLWQYTDAELAR  
 INQPLSAFASGSSSSSSSSSGTSSGAHAQLTNPLLASAGLSSNGENANPDIAHLCNVA  
 DHRLYKIVKWCKSLPLFKNISIDDQICLLINSWCELLLFSCCFRSIDTPGEIKMSQGR  
 KITLSQAKSNGLQTCIERMLNLTDLRLRLRVDREYEVAMKVIVLLQSDTTTELQEA VKV  
 RECQEKALQSLQAYTLAHYPDTPSKFGELLRLIPDLQRTCQLGKEMLTIKTRDGADFN  
 30 LLMELLRGEH

**36. SEQ ID NO: 36 Accession No. NM\_165364 Drosophila melanogaster**  
**Hormone receptor-like in 39 CG8676-PDHr39)**

35 1 actaacaaaa caaacatttt gctacttcgt cgcaggcggg actgtgttgc gtcgtgtgat  
 61 cgctagagcg gttgtggaat cggattcgag cgcaaaacac cgttcattgt gtgagcgaaa  
 121 aagagtggta gcgcctacag tggcatatgt agttaaatcc gtgaataagt gaaaaatccg  
 181 atatttgcg tgcaataatt tctcgtattg gcatcaagtg gcttccagtc gggtagatat  
 40 241 tgcacaagaa atgttatagc cataatgtgc acgcaaatta aacgaattct ctatgaaaat  
 301 gtgactagaa tgtgagtcga acaaaacgag taaaacgtga aatcccaact ggcttttggg  
 361 taacaaatct tatcaacaca gcaacggaaa tacattaaaa tcttgataga ctgagaaagg  
 421 gacaattgga atacttttag ttatttttaa atgttttaca acacaatgga actgcatcaa  
 481 cgacacctct caaactttta caaattgcac aactgagaaa tagtctttga taaataaata  
 45 541 aaatataaga aatcgctact gaaacaagat gccaaacatg tccagcatca aagcggagca  
 601 gcaaagcggg cctcttggag gaagtagcgg ctatcaagta ccggtcaaca tggcaccac  
 661 cacagtcgcg aatacgacga ccactttggg aagctccgcc gggggagcca ctggctcccg  
 721 gcacaacgtc tccgtgacaa acatcaagtg cgaactagac gaactaccgt caccgaacgg  
 781 caacatggtg ccggttatcg caaactacgt tcacggtagc ttgcgcattc cactcagtgg  
 50 841 acattcaaat catagggagt ccgattcgga ggaggagctg gcaagtattg agaactgaa  
 901 ggttcggcga aggacggcgg cggacaaaaa tggctctcgt ccaatgtcct gggagggcga  
 961 gctgagcga actgaggtca acgggggcga agagctgatg gaaatggagc caacaattaa  
 1021 gagtgaggtg gtccctgctg ttgcaccccc acaaccgctc tgcgcactac aaccgataaa  
 1081 aacagagcta gagaacattg caggcgagat gcagattcaa gagaagtgtt acccccagtc  
 55 1141 caacacacaa catcacgtg ccacaaaatt aaaagtggcc ccgacgcaa gtgatccgat

1201 caatctcaag ttcgaaccgc ctctgggaga caattctccg ctactggctg cacgtagcaa  
 1261 gtccagcagt ggaggccacc taccactgcc aacgaatccc agtcccgact ccgccataca  
 1321 ttccgtctac acgcacagct cccctctgca gtcgcctctg acgtcgcgc acccccccta  
 1381 cactccgtct ctgagccgca acaacagcga cgcctcgcac agtagctgct acagctatag  
 5 1441 ctccgaattc agtcccacac actcgcccat tcaagcgct catgccccac ccgccggcac  
 1501 gctctatggc aaccaccatg gtattaccg ccagatgaag gtggaagcct catccactgt  
 1561 gccgtccagt gggcaggagg cgcagaacct gagtatggac tctgcctcta gcaatctgga  
 1621 tacagtgggc ttaggatctt cgcacccgc atctcggcg ggcatatcac gtcagcagtt  
 1681 gatcaactcg ccctgcccc tctgcggtga caagatcagc ggatttcatt acgggatttt  
 10 1741 ctctgcgag tcttgcaagg gcttcttcaa gcgcaccgtg caaaatcgca agaactacgt  
 1801 gtgcgtgct ggtggaccat gtcaggtcag cattccacg cgcaagaaat gtccagcctg  
 1861 ccgcttcgag aagtgtctgc agaagggaat gaaactagaa gcgattcggg aggaccgaac  
 1921 ccgtggcggc cgtccacat accagtgtc ctacacgctg cccaactcaa tgetttagtcc  
 1981 gctgcttagt cctgatcaag cggcagcagc tgccgcccga gcagcagtg caagtcagca  
 15 2041 gcagccgcac cagcgactac atcaactaaa tggatttga ggtgtacca ttccctgtc  
 2101 tacttctct ccagccagcc ctagtttggc aggaacttc gtaagtcgg aagagatggc  
 2161 ggagacgggc aagcaaagcc tccgaacggg aagcgtacca cactactgc aggaaatcat  
 2221 ggatgtagag catctgtggc agtacaccga tgcagagctg gcccgcatca accaaccact  
 2281 gtccgcattc gcctctggca gctcttcgtc gtcgtcatc tcaggtacat cctcaggcgc  
 20 2341 ccatgcacaa ctaccaatc cactactggc tagtgctggt ctctcgtcca atggcgagaa  
 2401 tgccaatcct gatcttatc ctatctctg caacgtggct gatcccgct ttataaaat  
 2461 cgtcaaatgg tgcaagagct tgccgcttt taagaacatt tcgatcgtg accaaatctg  
 2521 cttgctcatt aactcgtggt gcgagctgt gctctctcc tgcgtttta gatcaattga  
 2581 tactcttga gagattaaa tgcacaagg caggaagata accctatgc aggccaaatc  
 25 2641 aaatggcttg cagacttgca ttgaacggat gtcaaccta acagatcac tgaggcgatt  
 2701 gcgcgttgat cgctacgaat atgttccat gaaagtatt gtgctgttc agtcagatac  
 2761 gacagagtta caggaagcgg taaaggtgcg cgagtgtcag gaaaagctt tgcagagctt  
 2821 gcaagcttac accctggcgc attatctga cagccatcc aagtttggg agcttttgc  
 2881 acgcattcct gatttgacg gaacgtgcca gcttggaag gagatgtga cgtcaagac  
 30 2941 tcgcatgga gctgattca atttgcta gtagctttg cgcggagagc attgacaatt  
 3001 gataactaag acggaaatct ttaccattg gcaaaacaag ttacatat ttagtattag  
 3061 atatatatat tctatagata agatccttac tgtaagtct gaaaacatgt gcctaaaaac  
 3121 caaagccacg atagcagtc catcaggccc actggctgag attaaatcca agagcaagat  
 3181 tgccaaattt ttacaccaat atatatttg atatgagcca tgtgcagggc ctcatcgc  
 35 3241 tgttgtgtc ggctaaagt tcagtaagaa aagtatatat tgattttgct atttatacat  
 3301 atttgactta tgtatagtgt aaactaaagc acacatggaa aatgaaaaga ctaaacaaat  
 3361 ttatttaaag attacttta ctattataga aaaaggggaa aaataaaaaa cacaaggca  
 3421 gagaagaaaa ttagttaca acaggtagcg acattttat attttctat ataaggaaat  
 3481 attcaatgta tttaaatat aaagccaaac ccgatttgg ttgggaaaga gctactgaaa  
 40 3541 ttttgatat ctatatatt atcactagaa gacgaatgaa tgtatccaat gtttaaattg  
 3601 tgtagcgtt agttttagt caatttcaca catgtctaca tacatgaata ttcagcgaga  
 3661 tatgtttgca aactattata aagcaaaaga cactcgaaa tcgccatcac tgggttggt  
 3721 aagactattc cagttatgct gttgttgca taaaaacca caactacgta catcaataaa  
 3781 atgtataatt tttattgga gtttagatt tgtattaact tctcctat aattacgatt  
 45 3841 attattatla ttactaatt tatgaatatt gtgtaacact gacttaaata gctgaaaaaa  
 3901 tctgcaaca ggatttaaaa cacctgaata caaaaacat tataacatga atacatttg  
 3961 ctatggcct agatagttg atatgtact tgcatatgta tgcagtgtc tatatgtgag  
 4021 tacgtaccat acaattcct gtcccaccag aaaaatcaca cgcaataaaa aattccaaaa  
 4081 tactaagtc gtatctaca agaaagatta aaagacaaat tgatgaatag gaatatgtg  
 50 4141 ccggaagtcc aagagattg gctgaaagta tcgacaaatt tcaacacat cgttcatgga  
 4201 tattgtgcta acactctcag ttgaaaatc atttctgtt aaactttta tataataagt  
 4261 tctccattcg attttgatt tacaattgt tctttaatt ttctttatc agttgtatc  
 4321 atgaaacatg aggatctcag tcatattga tctgttct ctgccgtaca ccgttctgt  
 4381 ccgttaattg aaaccataag tataatgaa attagttaa tgtttatta taaataaagc  
 55 4441 gctataataa attcaatac attatcata gtaactgat taagaccact gaaatcaaaa  
 4501 atatattat tactaagcaa agcacacgca aacaattat aatgttatt acgttaacaa  
 4561 caaactcatt ttaataatt ctatgaat acacaaagt acgcaattt cctctagc  
 4621 gcattgctta aatagttaa gaaaaaat aaacccatag cgcaatatt aatgtaaaac

4681 agttttcctt gcgtgtgatg ttgctctag ctacgtacaa attcatcatt tattaaattt  
4741 aaaactcaat ttgctttta aataaattta ataagtaaaa ttcaacaata attgatatac  
4801 aattgtcaat gcaatatttt gtaataaaaa tgcgaaaaat c

## 5 37. &gt;SEQ ID NO:37 -- 96\_Æ\_Ex4\_7.55\_kb+oligos\_Map.seq

GGGCCCCCCTCGAGGTGACGGTATCGATAAGCTTGCCGGTGGCGGAGAAGGTGTATCCGTGATTAAGAAAGAGCCAGC  
CGATGAGAAGCAGCCACAGCCACATGACCACGGTGGCTCCGTACACAAGATGGCACTGCTGGTGGCGTTTCGAGACCGAT  
TTGAGGAACTCCTCCAGTTCGTCCCCACATGACCGCTTTTGAAGCGGCAGGGCGTGGCGCACCACATCTTTGTGCTG  
AACCAGGTGGACAGGTTCCGCTTCAATCGCGCTCTCTCATCAACGTGGGTTTCCAGTTTGCCAGCGATGTGTACGATTA  
10 CATTGCCATGCACGACGTAGACTTGCTGCCCTTGAATGACAATCTGCTCTATGAGTATCCCAGCAGCTTGGGACCACTGC  
ACATCGCCGGACCGAAGCTACATCCCAAATACCACTATGATAACTTCGTTGGAGGAATATTACTGGTGGCAGCGAGCAC  
TTAAGCAGATGAACGGCATGTGCAACCAGTACTGGGGCTGGGGATTAGAGGACGACGAGTTCTTCGTGCGCATCCGGGA  
TGCAGGACTGCAGGTGACGCGGCCGAGAACATTAAGACTGGCACTAATGATACATTGAGGTGAGACCAGTGTCTCCGGAT  
TTCGCAACTAGACGTGACTACTAATAATTATTGTCAATCAACCTCAGCCATATTACCAACCGCTATCATCGTAAGCGGGA  
15 CACCCAGAAGTGCTTCAACCAGAAGGAGATGACCCGCAAGCGGGACCACAAGACGGGCCTGGACAACGTGAAGTACAAAA  
TACTTAAGGTGCATGAGATGCTCATTGACCAGGTGCCGGTGACCATCCTCAACATTTTGCTCGATTGTGATGTTAATAAA  
ACGCCTTGGTGGACTGCTCCGGAACGGCAGCGGCTGCATCGGCGGTACAAACCTGATGGGTTGTGTTAAACCAAGATC  
CTATGTTTATTTTCGCTATTATAGTGTGTGTTGTTATGTTATAAATGCGCTAATACACGTGCACCATGCCATAGAGGAATGTCC  
AGAAGAGCACGTAGGTGCAAAGGCCGCCATGAACTGATTGGTFCAGCAGATTTCTGCGGTTAATGAAAAACTTGCGCCAC  
20 TGGGTGCCCCGATTTACGAGCACCAGAATCCAGAGCACGAACACGGACAGGAAGTAGAAAAGGAATCCAGCGTACCACT  
CAGGCCCAAAATACCTGCGAATTGGTGGGACATTAAGTTGGTTTACCATCAATTGGAGCCAATTACCCGCAGCGCA  
GCCCCGAGATGGCAGCCATCGATGTGCGACAGTATTCACGGCGGATATGTTGTTCCGGATGGCGCCCTCGCTGTAGGCGA  
TTATTTCCGCCAGTCTTGGACTGCGTGGTCTTCACTCGATTCAATTTATTTAATTAAATTTCTACTTTAATTTCTAGCAAAA  
ATATTCCTAGGCTGTGAACTTCGATTGTGTGCCGATTGTGTTATCGATTGGTGGCGATAACTATGCACTGTAAAAATTCA  
25 CTAGCGGTTTTTGCAGGATAAATAGTTTTTGTAAATTTTCCGAGATAAACTTGACGAGCTGTTAATGTTAAATAATGAA  
GTTTAATACAATATCAAATATATTTGCTGAAGTGTATATTTATTTCTCACCCTCTGTGCTTCGATGGCTCACAATTGCGT  
TTGCCATTCCGCCGGGCACGTAGATTGTTGTTATTGGGATTGGCCTGGAGCACTCGGACGGACAGTAATTCATTAAATA  
TGTGGTGATAACGCGAGCTGCCGAATCTGCGTGCAATTCGTGCGTTTGACGTGGGTACTAACTGCTATGCTGTGCGCGCG  
ACAGTTGTTCTGATACGCAGAGTTCCTGCCTCACCACACACGACCACCTCCATTAAACCAGCCACCCCCCCCAGCGCCT  
30 CCTCCACCGACAGCAGCTGCTCCACCGCACCACCAGGAGAGGGGCAATTAATAAATCAATCAGAGGGCCCATCACTTGCT  
TGTAACCGCCGAAGAACTGCGCGGTGTGCGGGGACAAGGCTCTGGGCTACAACCTTCAATGCGGTACCTGCGAGAGCTGC  
AAGGCGTTCTTCCGACGGAACGCGCTGGCCAAGAAGCAGTTCACCTGCCCTTCAACCAAACTGCGACATCACTGTGGT  
CACTCGACGCTTCTGCCAGAAATGCCGCTGCGCAAGTGCTTGATATCGGGATGAAGAGTGAAAACATTATGTCCGAGG  
AGGACAAGCTGATCAAGCGGCGCAAGATCGAGACCAACCGGGCCAAGCGACGCCTCATGGAGAACGGCACGGATGCGTGC  
35 GACGCGGATGGCGGCGAGGAAAGGGATCACAAAGCGCCGGCGGATAGCAGCAGCAGCAACCTTGACCACTACTCGGGGTC  
ACAGGACTCGCAGAGCTGCGGCTCGGCGGACAGCGGGGCAATGGGTGCTCCGGCAGACAGGCCAGTTCCGCCGGGCACAC  
AGGTCAATCCGCTTCAGATGACGGCCGAGAAGATAGTGCACGAGATCGTATCCGACCCGGATCGAGCCTCGCAGGCCATC  
AACCAGTTGATGCGCACGCAGAAAGAGGCTATATCGGTGATGGAGAAGGTAATCAGCTCACAAAGGACGCCTTAAGGCT  
GGTGTGCGATTTGATCGACTATCCAGGTGGGTGCAGACAAGATTTTCATCGTTTAGCCTTATCCGCTCACCTATGAACGAC  
40 TTGAATCTTTACAGGCGACGCACTCAAGATCATTTCAAAGTTTATGAACTCGCCCTTTAACGCGCTGACAGGTTAGAGTT  
TTAAATTTGTGGTTTTAACTTAATTTACATTCCTTGTAAATTTAAATACGCACTATTACCAAATTCATGAGCTCAC  
CCACGGACGGCGTTGAAATTTATCTCAAAGATAGTTGATTGCCCCGCGGACGTGGTGGAGTTTATGCAGAACTTGATGCAC  
TCGCCAGAGGACGCCATCGATATAATGAACAAGTTTATGAATACCCAGCGGAGGCGCTGCGCATTTCTTAACCGAATCCT  
AAGCGGCGGAGGAGCGAAGCGCAGCCAGCAGACAGCAGACCGCAAGCCATTGCTGGACAAGGAGCCGGCGGTGAAGCCTG  
45 CAGCGCCAGCGGAGCGAGCTGATACTGTCAATCAAAGCATGCTGGGCAACAGTCCGCCAATTTCCGCCACATGATGCTGCC  
GTGGATCTGCAGTACCACTCGCCCGGTGTGCGGGGAGCAGCCAGTACATCGAGTAGCCACCCCTTGCTTACATAGCCAA  
CTCGCCGACTTTCGATCTGAAGACCTTCATGCAGACCACTACAACGACGAGCCAGTCTGGACAGTGATTTTAGCATTA  
ACTCAATCGAATCGGTGCTATCCGAGGTGATCCGCATTGAGTACCAGGCCCTTCAATAGCATAACAAGCGGCATCGCGC  
GTAAAGGAGGAGATGTCTACGGCACTCAGTCTACGTACGGTGGATGCAATTCGGCTGCAAACAATAGCCAGCCGCACCT  
50 GCAGCAACCCATCTGCGCCCCATCCACCCAGCAGTTGGATCGCGAGCTAAACGAGGCGGAGCAAATGAAGCTGCGGGAGC  
TGCGACTGGCCAGCGAGGCTCTTTATGATCCCGTGGACGAGGACCTCAGCGCCCTGATGATGGGCGATGATCGCATTAAG  
GTAACCCGCTAGGGATAACAGGGTAATAACAGTCCACGGTATTAGCCTATAGGTCTTTCTACATTTATAGCTCCAACACC  
ACGGCTTATCTAATCAGAGTGTGCGAGCTGCGATATATGTACACACGGCACCTGGCACTTTTATAGCCATTCGGTGATTCA  
GTGCGTCTCTCGATGTTGGCCACGGGCCGTATCTTCGTGAGCCAGTTTCTGGGTTCCAGCAATGCTCGCCTACCAAT  
55 GTAAACACACTTTTTAATGGGGTGGCTCAAAGTTTTTGATTTCCCAAGAGCTTTGGTTCGAGTAAAAGAAAATTGATCGAA  
CCAGATAAGCTATTTTCCCCCAGAGGGTTAAAGAATTTGAAGTCATGCGACTGGGTCTAGTTAAGATATTTGATTACGAA  
AATTGGCCTTTAATTAAAGACCCTAAACGTGACAACTTCCATTCTATATACTTCTTGATGAGTATTTAAACAAATATGGC  
TATTTTCGGAACAAATCGGGCACTCATTTATATCTTTAGCTTTATCTTTATTTTAAAGATGTGTCCACACCTTTGATCG  
ACCTCTAGTTCCCTTGAGAAATGATTTGGAATTATCCAATAATGATTCATCACTTCCACGAATTGTTGTCCCATTAAATC  
60 GAGCCACCTAGCTTTTCATGCAATCAGAAGCTCTGGTCTGCCAAGAAGGAGCAGACAGCGCTTTATCAGCCTCTGGGCG  
TGCCAATTGTGACACTATCAACCATTATCAAGAGTACCAGCAGGCGCTCATGAGTCTCCAGCCAGCCGTGATTTGGGCG  
TTTATTTGTCGCTTAGACTGTTTACCGATTTTGCCTCATCGCAATTAGCACATTTAGTATTGTTAATTGGGAAAAACGAT  
ACAATTTTGACGAAATATATGGAGCAGCCAGGTGTTGGGCGCTATGATAAGCAGTGCTCCGCCATTCGATTGAGTCACCT  
TCCAGGGAGAAGCCTTTACGATTATGGCGATAATAATGGCCACCAAGAGAACATGGGCAACATACGCACTGACCTGCTC  
65 AAGTTTGGCGAAGGCAATATCTACGAGGAGCACCAAAAGTTCATCACACGTTTGACGAGAAGTGGCGCATGGACGAGAA  
CATAATCCTGATCATGTGTGCCATTGTCTTTTACCTCGGCTCGATCGCGAGTGATACACAAAGACGTGATTAGATTGG  
AACAGGTGAGTAAGCACTTGATACCATACTGCAGTATTACTAATTTCTTTTATTCGATAGAAATCCTACTATTATCTTC  
TGCGAAGATATCTGGAGAGTGTATTCTGGCTGTGAGGCGAGAAACGCGTTTATCAAGCTAATCCAAAAGATTTGAGAT  
GTGGAGCGTCTGAACAAGTTCATAATTAATGTCTATTTGAATGTTAACCATCCAGGTGGAGCCCTTGCTGCGTGAAAT  
70 ATTCGATTGAAAAATCACTAGACAACCGATGCGTGTGCGGCATTTAATGCCTATGTTGATGCCCAATGATGAATGGTCA  
ACAAGCTGTAGTTGTTGTTGTTGTTGATGTCTGTTTATCTTGTCGCTTGTAATGTTAGATTTTAATCGAATGTGATTGT

TAGATTTGCATATACTGCATAGATTTTATATTTCTACATCAAAGAGAGCATATTTAGGATACCAAGTGCAAAGCAACACA  
ATCTATATGTAATGTACACCGTTTACCTAGTTTCAAATAAACTAGACGATAATGCAATAACTTAAGTTGGAAGCGTGGGTT  
CTGTGCAAAAAGGAAAAAGACAAAAAAATAAACTGACTTTGAGAACCAGTGGTAATAAAATGTCTCGTATTCTTTTCT  
ACTCGAATGAATTTTGAACCCCTCCAGGACAAATTACGCAACGAGTGATTTTGAACAACAATCCAAAATAATTTAATTCC  
5 GAAAGTCACAAAATAAAAATTCTGAAGTAGGAAAAACAAATAAGATGTTTGGAAACCAACGAGAGATGTGCTTCGTAA  
GCATCAACCCGGGGAAACACCACAGCAACCGCGCATGTGTACCCGCGACCAGTCTTCAGAAATCCACGTCGTGTACGTAT  
CCGCGAGCCAGCGTATGTGTCCGCATCTGCCGACCCCGTCTTACATAGTCATTTATGTATAATGTAGGTAATATAATAGCT  
CGAGCTCGCTCCGCACCACCAATGTGCGTCGTGCAAGTCCATTCCAATTGTTATCCGGTCCACTCGCCGCGCAAATCGGC  
TTTCAGGTTGATTCGCGGCAATCCTTGGCCCATTTGCAGAACTCATCCAACGCGCTGACGGCCAAATTGCGAGAAAGAGC  
10 CTTACACGCGAGATTACGATCGGTTGTAATGAGCACCAATTCGTTTGAATGAAACACTTGCCATCTGCAAAAGAGTTTT  
AGTTAGAAATGCTATCAGGAAGGACATTTAACGGAAGCAGCTCACCTGTGCATTGTTTCGTTTTCGCGTTTTCGAGACA  
GCCATTGCCGTTGCCAGAATCTTGTGTCATTGGACAAATATTCCTCCTCAACTAGGGCAAACACCGATGCATTGACAAA  
CGAGCCCTTTGTGGTAGCACATCTAGAAAAGAAATCAATAAGGTATTATTGATCAGCAGGAAAAGCTTTCCTGAACAACT  
15 ATTACTACTGATTTAAAAGTAAAATTTCAATACATTATCAGGAACTTTTATCTATCTCAATAGCAACCAATGAATTAGA  
CAGAATTATAAATAGCTAATCGCTAGTAAACCTTTATCAGATATCAGTAATAAAGGAACATGAGCTGACGCGCGGAAT  
ATAATTAACAATAGCTTACTTCACATTGCCTTTGGCCGACTTGATGAACTCTAACGACTTTTTGGCCCGCGACGACACCT  
CGTCAAAGTGGTGGATGCGCTGCGTCTGCTTCGAACTGCGGTACGAGTCCAACCTTAACGCCCTTGGAGAGGCCATCCAGT  
TCCTTAACCACTGTCAGTGGTATAATAAGTGTGTAGCGTTTAACTCCGTGGACAGTTTTTCAAAGTCTTCAAGGCAGTC  
20 GATAAAGCAGTTGGTATCCGGTAGAAGATAGCGCGGTGCGACCTCGATGTATATTTTCGTGTCCACGAACCTTGAGAATGT  
CCTCCAACCTTGCTGGTGCATATCTGCTTAACCTAGCCTTGGCCTCAAGCTCTTCTTCAGCTTGACAGCTCGGAAACA  
TCGGAATCAGTTGAACGACACAACAATTCCTCAACGGCTTTACATTGAATCTTTGAAACGTTTCGCCGCAAGACCACAAC  
TTCAGCATCATAGTTTTCCAATGCGTTCTTCATCGCTGCGTTTAGGTGCTCGACAGTGAAGCTTGATTCCAACGGCTGCT  
GGAGAGCCTCTTGATGCTGGACGTAGTATTCTGGAACCTGACCAATCCTACGAACACGTTGCAAAAACCTGAAGACGTTG  
25 GATCCTCTTTTGAGATAGTCCATGCTCAGCGTTTCACGACCAACGGAGTGAAACCTCTAAGGGCCACATCCTCATCAA  
CAGTATTTCTGTTTTCTCAAGTTTGTGCTTCCCATTAAGCATTCAATGTACTCAAAAAGGATTGTTAGCTCAGCCCAGC  
AATCGATGAAAGAGTGTCTGCAATATTGGAGTTAATGAAAACTAATAAAGGCATTCAATTTATACATACTCTTCCGAT  
CTGACGGGTTCCACACGTCCAAACCTGATGCTCAACCAACGTACATAAACATTACGAACTGTAGGTATGTGTTACAGT  
CGCAAAGCTGAAATAAAGATTAATTAGCAATAATAAATAACAAGGCGAATTTTAGCTTACTCTTCTTCTGGGAGACAC  
TGGACATTTGTAGAATCCTCTAGATCTACTAGTCC

**38. SEQ ID NO:38 >GAL4-DHR96\_DNA\_**

GAAGCAAGCCTCTaGAAAGATGAAGCTACTGTCTTCTATCGAACAAGCATGCGATATTTGCCGACT  
TAAAAAGCTCAAGTTTCGcgatggcggcgaggaaagggatcacaaagcgccggcgatagcagcagcagcaacctgaccactactcggcag  
aaagaggctatatcggtgatggagaaggtaatcagctcacaaaaggacgcctaacagaggacgccatcgatataatgaacaagltcatgaataccccagctcgc  
35 ccggtgtcggggagcagccagctacattctacgtacgggtgatgcaatctgaagttcatcacaaagttgacgagaagtggcgcatggacgagaacataatcctg  
atcatgtgtgccattgtcctttaatgtctatttgaatgttaacccatcccaggtggagcccttgctgctgaaatattcgatcaaagagagcatatttaggataccaagtg  
caaagcaacacaatctataagacgataatgcaataactaactggaagcgtgggttctgtgcaaacc

**39. SEQ ID NO:39 >pET24c\_Bam+Xho\_filled+DHR96**

TGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAAGCGCGGCGGGTGTGGTGGTTACGCGCAGC  
GTGACCGCTACACTTGTTAGGGTGATGGTTCTTAATACAACCTATTAATTTCCCCTCGTCAAAAAT  
AAGGTTATCAAGTGAGAAATCACCATGAGTGACGACTAACC GGCGCAGGAACACTGCCAGCGCA  
TCAACAATATTTTACCTGAATCAGGATATGCTTCCCATAACAATCGATAGATTGTCGCACCTGATT  
GCCCCGACAGATCTTCTTGAGATCCTTTTTTTCTGCGCGTTGGCGATAAGTCGTGTCTTGGTAGTGA  
45 GCGAGGAAGCGGAAGAGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTACACAC  
CGCAGGGAGCTGCATGTGTCAGAGGTTTTACCGGTCATCACCGAAACGCGCGAGGCAGCTGCGG  
CGATGAAACGAGAGAGGATGCTCACGATACGGGTACTGATGATGAAACGGAACCGAAGACCA  
TTCATGTTGTTGCTCAGAAGATTCCGAATACCGCAAGCGCTCACTGTCTTCGGTATCGTCGTATCC  
CACTACCGAGATATCCGCACCAACGCGCAGCCCGGACTCGGTAATGGCGCGCATTGCCGAGACA  
50 GAACTTAATGGGCCCCGCTAACAGCGCGATTTGCTGGTGACCCAATGCGACCAGATCGCTTTACAG  
GCTTCGACGCGCGCTTCGTTCTACCATCGACACCACCACGCTTCACCACGCGGGAAACGGTCTGAT  
AAGAGACACCGGAAGGAGATGGCGCCCAACAGTCCCTCTAGAAATAAAACCTTGACCACTACTC  
GGGGTCACAGGACTCGCAGAGCTGCGGCTCGGCGGACAGCGGGGCCAATGGGTGCTCCGGCACC  
TTAAGGCTGGTGTGCGATTTGATCGACTATCCAGGCGACGCACTCAAGATCATTTCAAAGTTTAG  
55 CTGCGCATTTCTTAACCGAATCCTAAGCGGCGGAGGAGCGAACGCGAGCCAGCTACATAGCCAAC  
TCGCCGGACTTCGATCTGAAGACCTTCAAGCAACCCATCTGCGCCCCATCCACCCAGCATTCCGT  
GACAACTATATCCGGAT

**40. SEQ ID NO:40 F96Xma**

5'-GAGAGATGTGCTTCGTTAAAGCATCAACCC

**41. SEQ ID NO:41 R96SpeBgl**

5'-GGACTAGTAGATCTAGAGGATTCTACAAATGTCCAGTGTCTCCC

**42. SEQ ID NO:42 R96Int3**

5'-CCATTATTATCGCCATAATCGTAAAGG

**43. SEQ ID NO:43 R96EX3SCE**

5'-ATTACCCTGTTATCCCTAGCGGGTTACCTTAATGCGATCATCGCCC

**44. SEQ ID NO:44 R96endhind**

5'-GGAAAGCTTTTCCTGCTGATCAATAATACC

**45. SEQ ID NO:45 FAPA96**

5'-TGGGCCCATCACTTGCTTGTAACCGCCGAAGAACTGCGCGG

**46. SEQ ID NO:46 F96INT3SCE**

5' CGCTAGGGATAACAGGGTAATAACAGTCCACGGTATTAGCCTATAGG

**47. SEQ ID NO:47 F96EX5Int3**

5' CGATTATGGCGATAATAATGGCCAAAGAGAACATGGGCAACATACGC

**48. SEQ ID NO:48 FGALXB**

5'-GAAGCAAGCCTCTAGAAAGATGAAGC

**49. SEQ ID NO:49 RGAL96**

5'-CGTGCCGTTCTCCATCGATACAGTCAACTGTCTTTGACC

**50. SEQ ID NO:50 R96/936**

5'-GCCTGGATAGTCGATCAAATGCG

**51. SEQ ID NO:51 F96BEG**

5'-ATGGAGAACGGCACGGATGC

**52. SEQ ID NO:52 F96XBai**

5'-TACATTCTAGAGACCAACTACAACGACGAGCCCAGTCTGG

**53. SEQ ID NO:53 R96BspE1**

5'-CATTCATCCGGACATTAATTATGAACTTGTTTCAGACGCTCC

**54. SEQ ID NO:54 R96BspE2**

5'-GGGCATCAACTCCGGAATTAAATGCCCCGACACGCATCGG

**55. SEQ ID NO:55 RPAXCRE-AN**

5'-GTCTCACGACGTTTTGAACCCAGAAATCGAGCTCGCCCCGGGG

**56. SEQ ID NO:56 RPAXCRECO**

5'-CACGAATTCCAAACTGTCTCACGACGTTTTGAACCC

**57. SEQ ID NO:57 FPAXFSE-AN**

5'-GAGAGCTAGCATGCCGGCTAGATCTCGAGATCGGCCGGCCTAGG

**58. SEQ ID NO:58 FPAXPOLY**

5'-GAACTGCAGCTCGAGAGCTAGCATGCCGGC

**59. SEQ ID NO:59 F96ANhe**

5'-GGAGATATACATATGGCTAGCATGACTGGTGG

**60. SEQ ID NO:60 R96AHind**

5'-TGCTCGAAGCTTCGCAGAAGATAATAGTAGG

## V. CLAIMS

What is claimed is:

1. A composition comprising an inhibitor of DHR96 activity.
- 5 2. A composition comprising an inhibitor of DHR96 activity and a pesticide.
3. The composition of claim 2, wherein the pesticide is selected from the group comprising tebufenozide, DDT, and phenobarbital.
4. An insect comprising a gene, wherein the gene comprises a non-naturally occurring mutation of the DHR96 gene.
- 10 5. The insect of claim 4, wherein the mutant has a defect in activation with retention of dimerization ability of DHR96.
6. The insect of claim 4, wherein the mutant has a defect in activation without retention of dimerization ability of DHR96.
7. The insect of claim 4, wherein the insect fails to modulate genes in the xenobiotic  
15 pathway.
8. The method of claim 7, wherein the gene is in the cytochrome P450 family.
9. The method of claim 7, wherein the gene is in the carboxylesterases family.
10. The method of claim 7, wherein the gene is in the glutathione S-transferases family.
- 20 11. The method of claim 7, wherein the gene is in the UDP-glucuronosyltransferase family.
12. A method of enhancing the effect a pesticide has on an insect comprising administering to the insect an inhibitor of DHR96 activity.
13. The method of claim 12, wherein the pesticide and the inhibitor of DHR96  
25 activity are administered simultaneously.
14. The method of claim 12, wherein the inhibitor of DHR96 activity is administered before the pesticide.
15. The method of claim 12, wherein the pesticide is selected from the group comprising tebufenozide, DDT, or phenobarbital.

16. A method of identifying an inhibitor of DHR96 activity, comprising the steps of:

- a. testing compounds for inhibition activity of DHR96 and/or inhibition of xenobiotic activity; and
- b. comparing the activity of these compounds to known inhibitors of DHR96.

5

17. A method of identifying ligands for DHR96, comprising the steps of:

- a. creating a fusion product comprising a DNA binding domain, a DHR96 ligand binding domain (LBD), and a reporter gene;
- b. expressing the fusion protein of step a, wherein the fusion protein is expressed in the presence of an appropriate ligand; and
- c. detecting reporter gene product, wherein said reporter gene product indicates the presence of a ligand that binds DHR96.

10

18. A method of manufacturing a composition for inhibiting DHR96 activity, comprising admixing the inhibitor with a pesticide.

15

19. A composition produced by the method of claim 19.

1/16

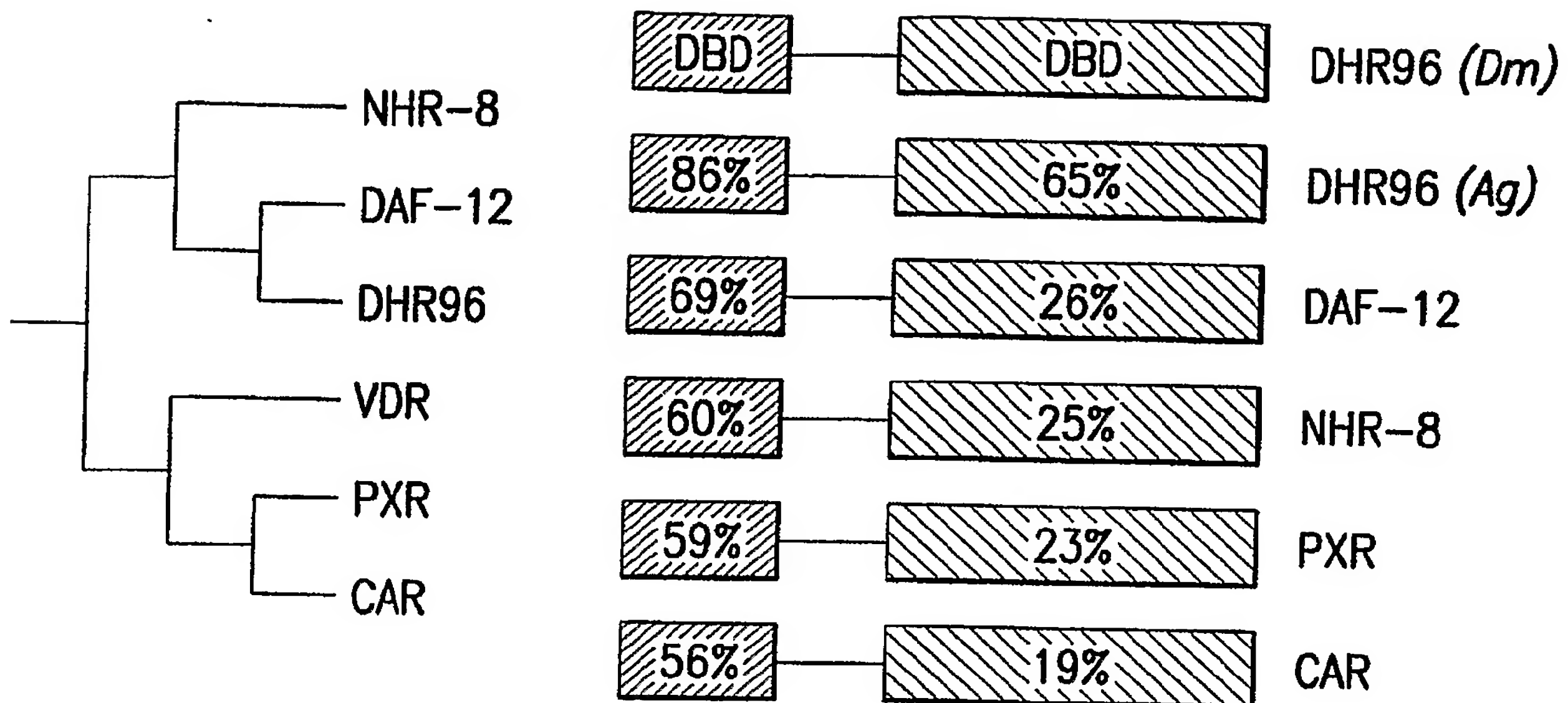


FIG. 1A

FIG. 1B

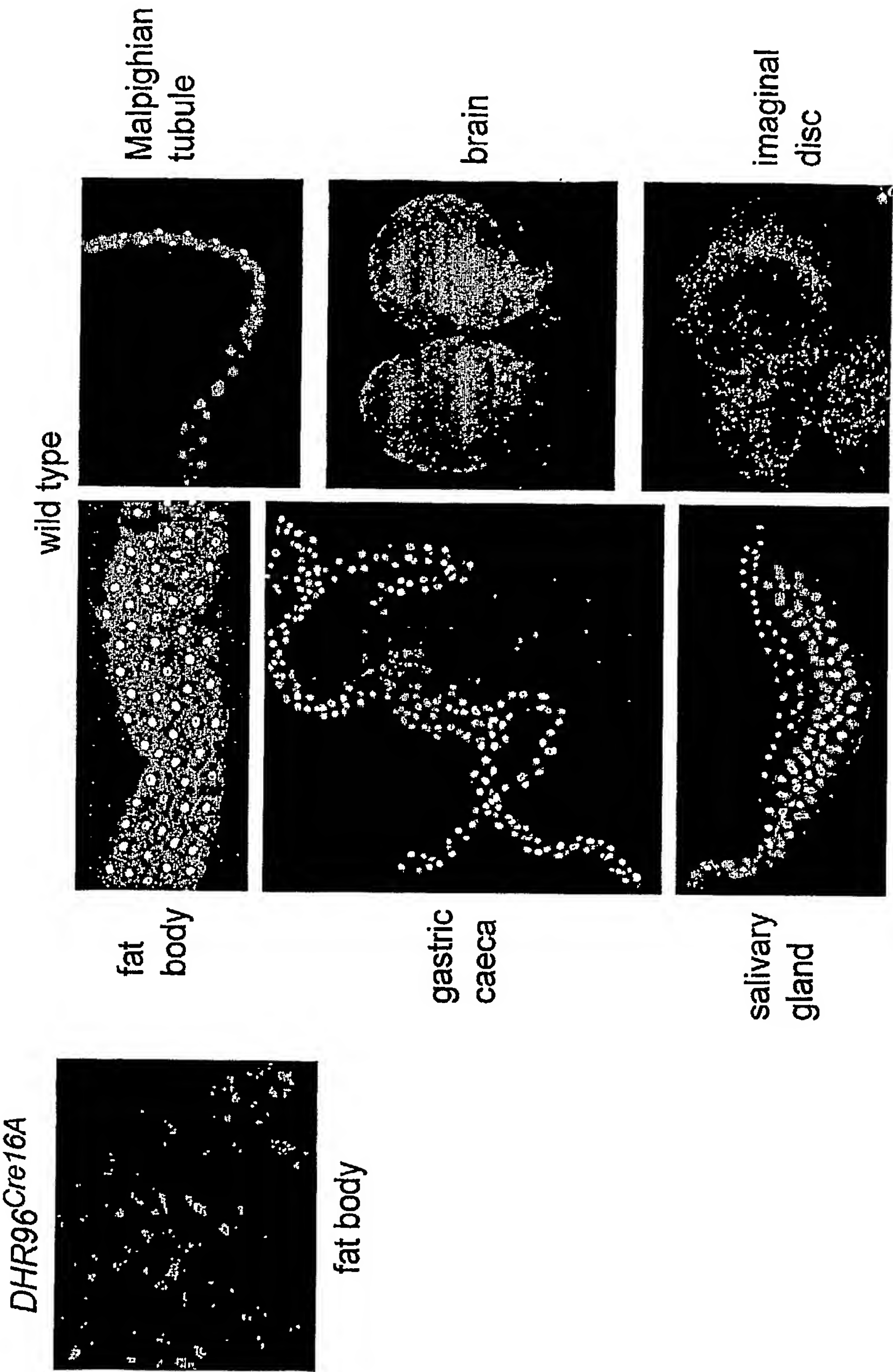
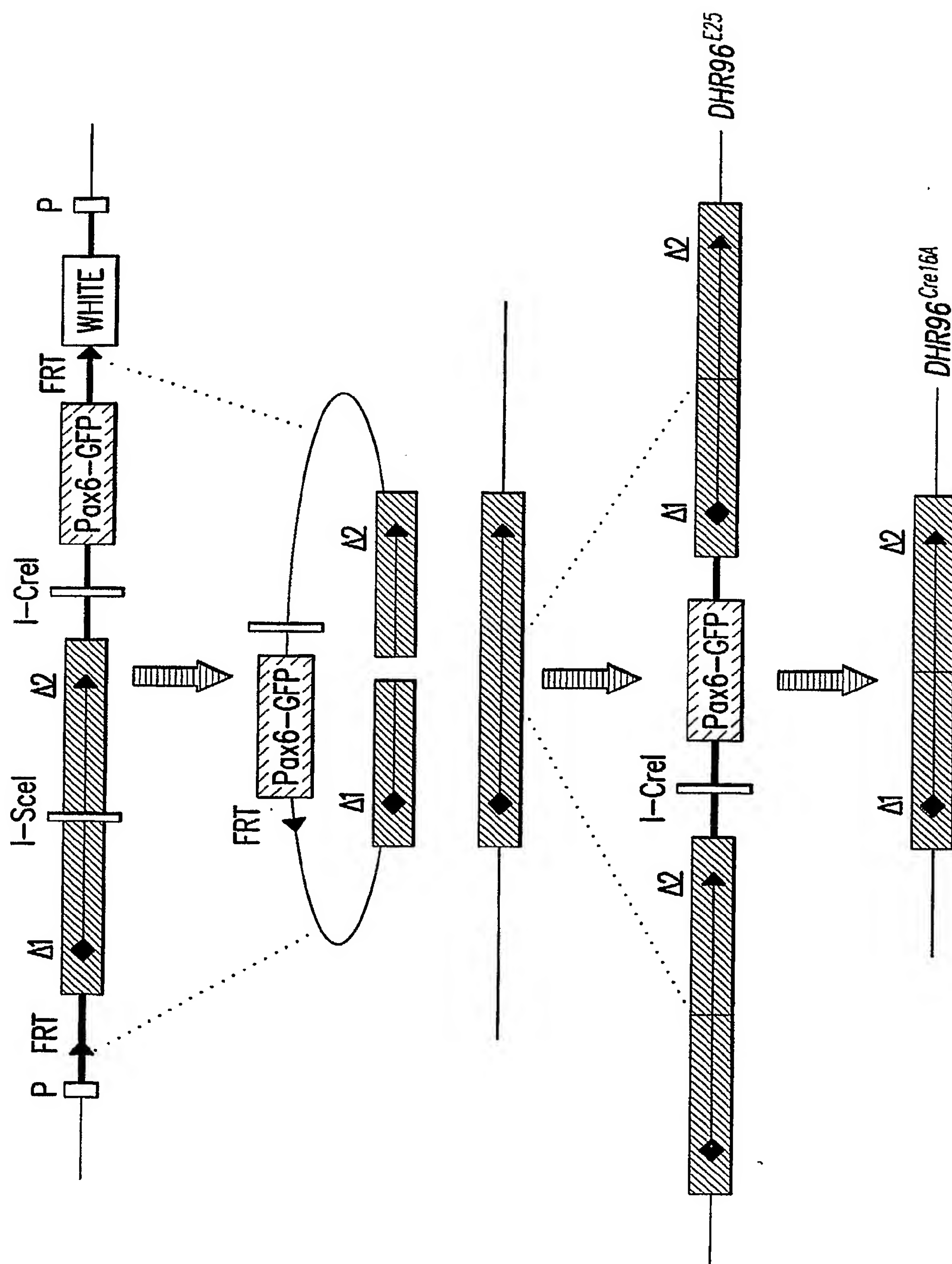


FIG.2



# FIG. 3

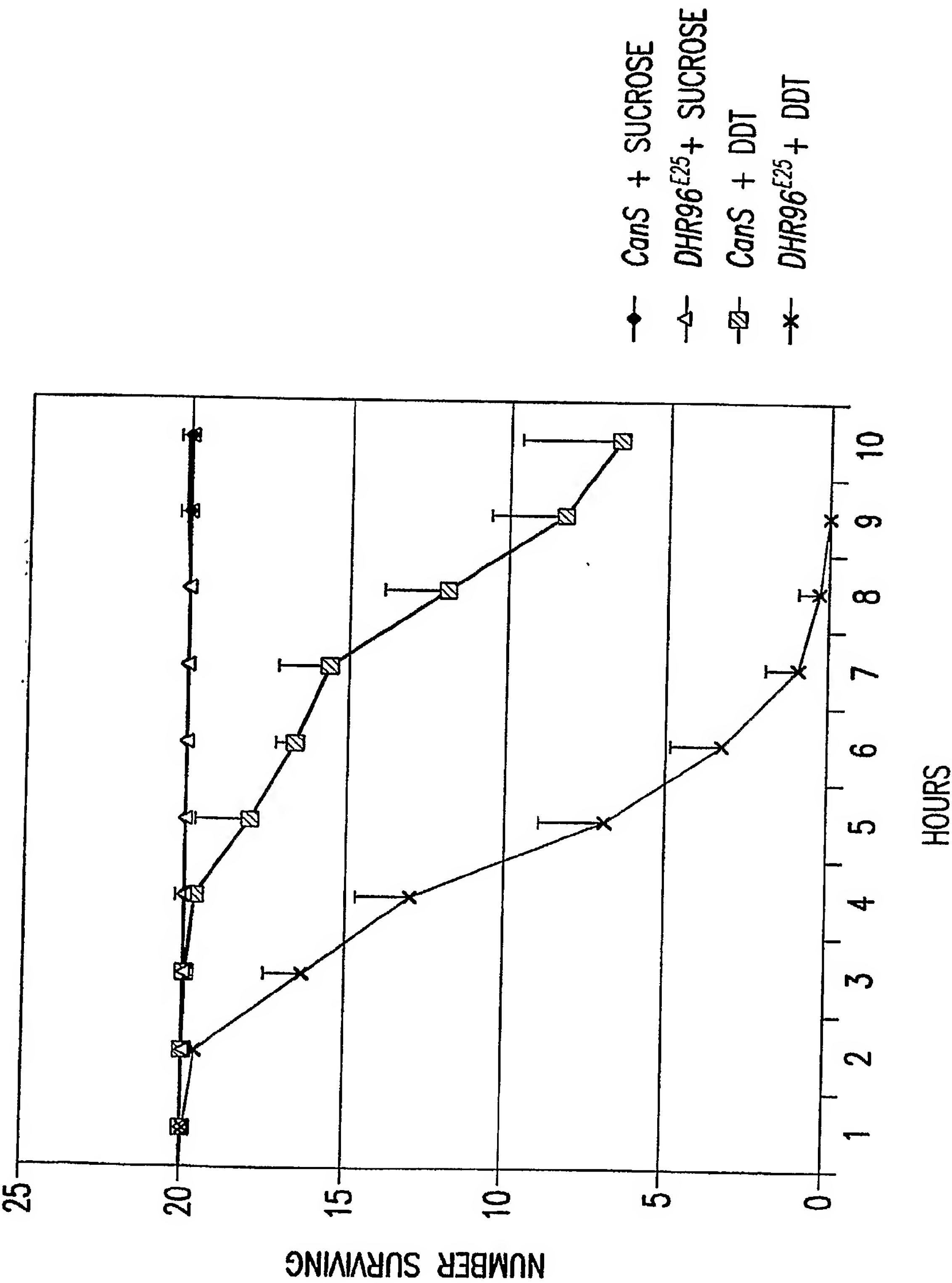


FIG.4

5/16

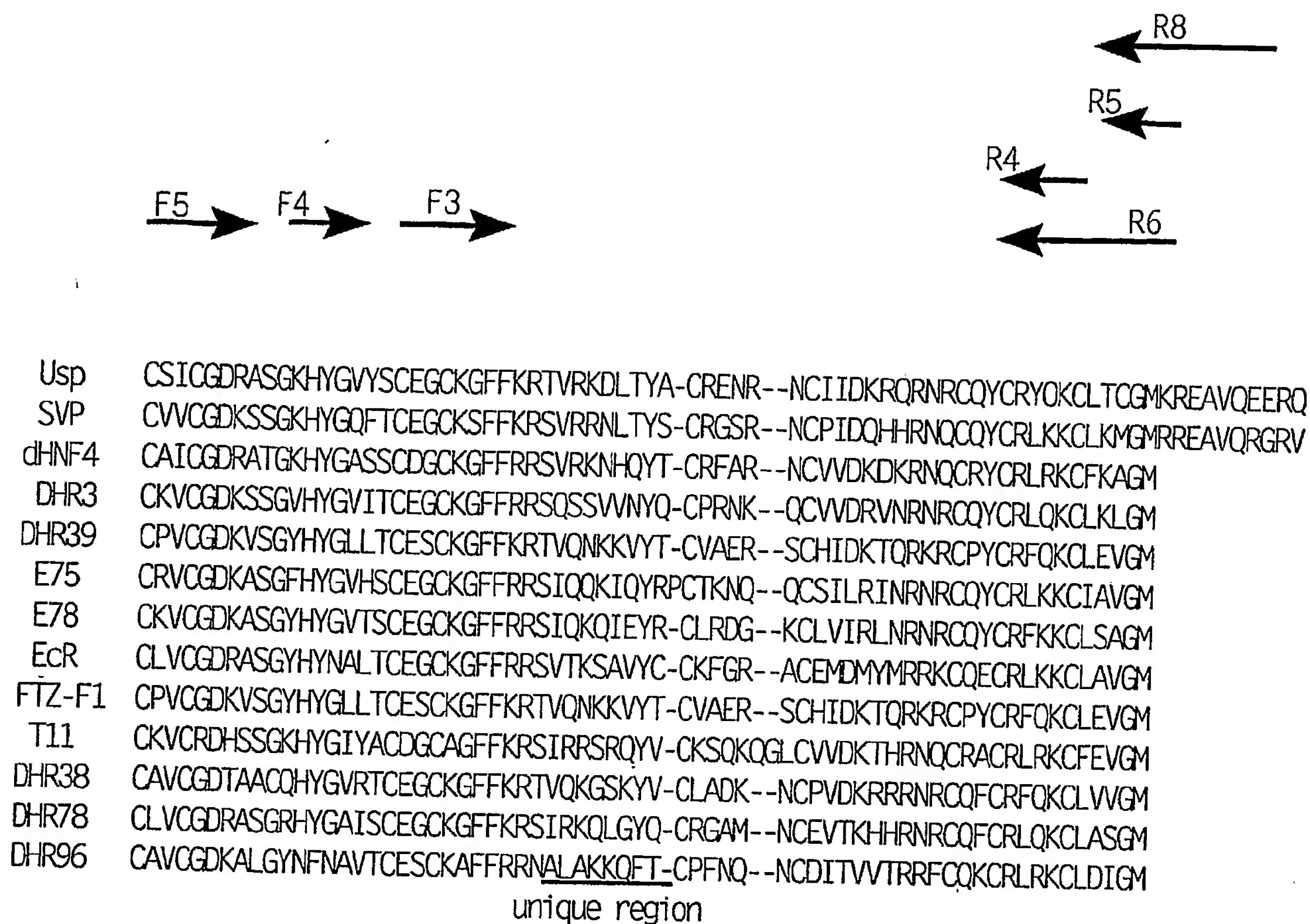


FIG.5

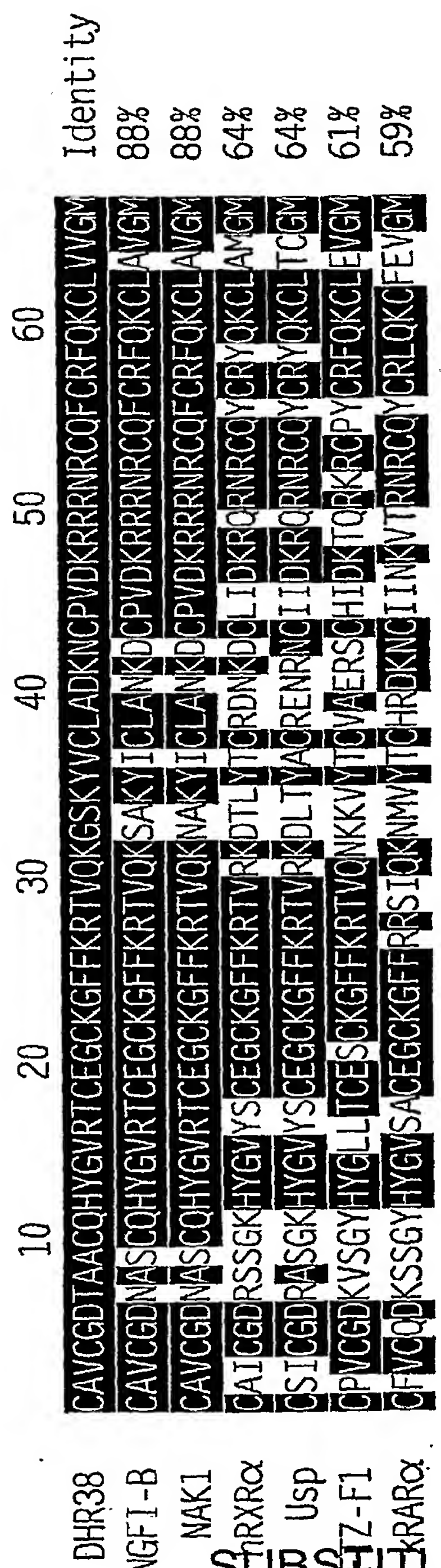


FIG. 6A

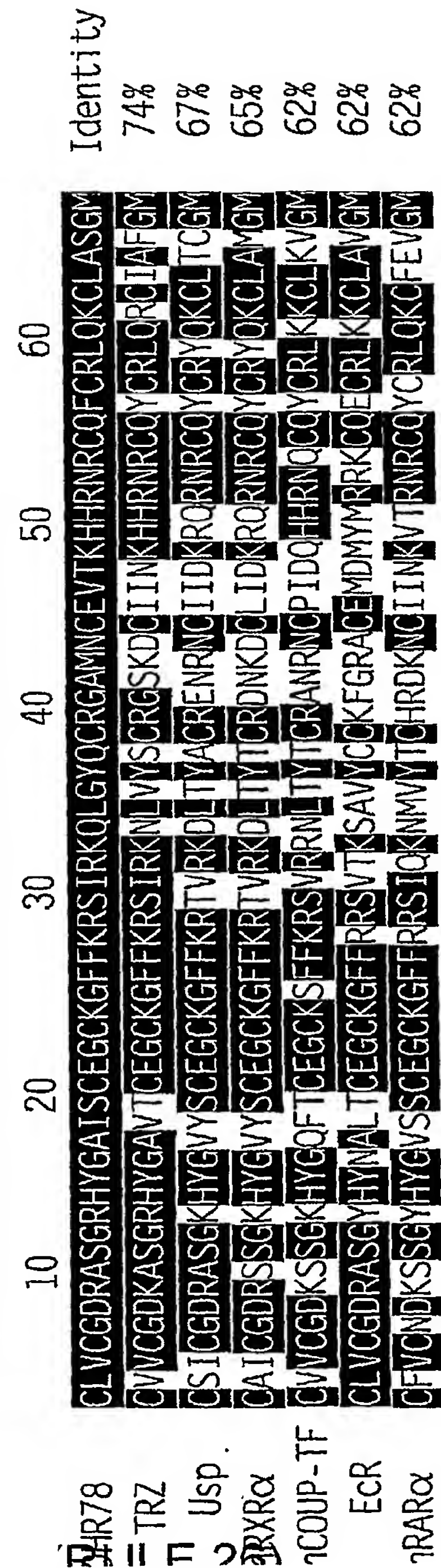
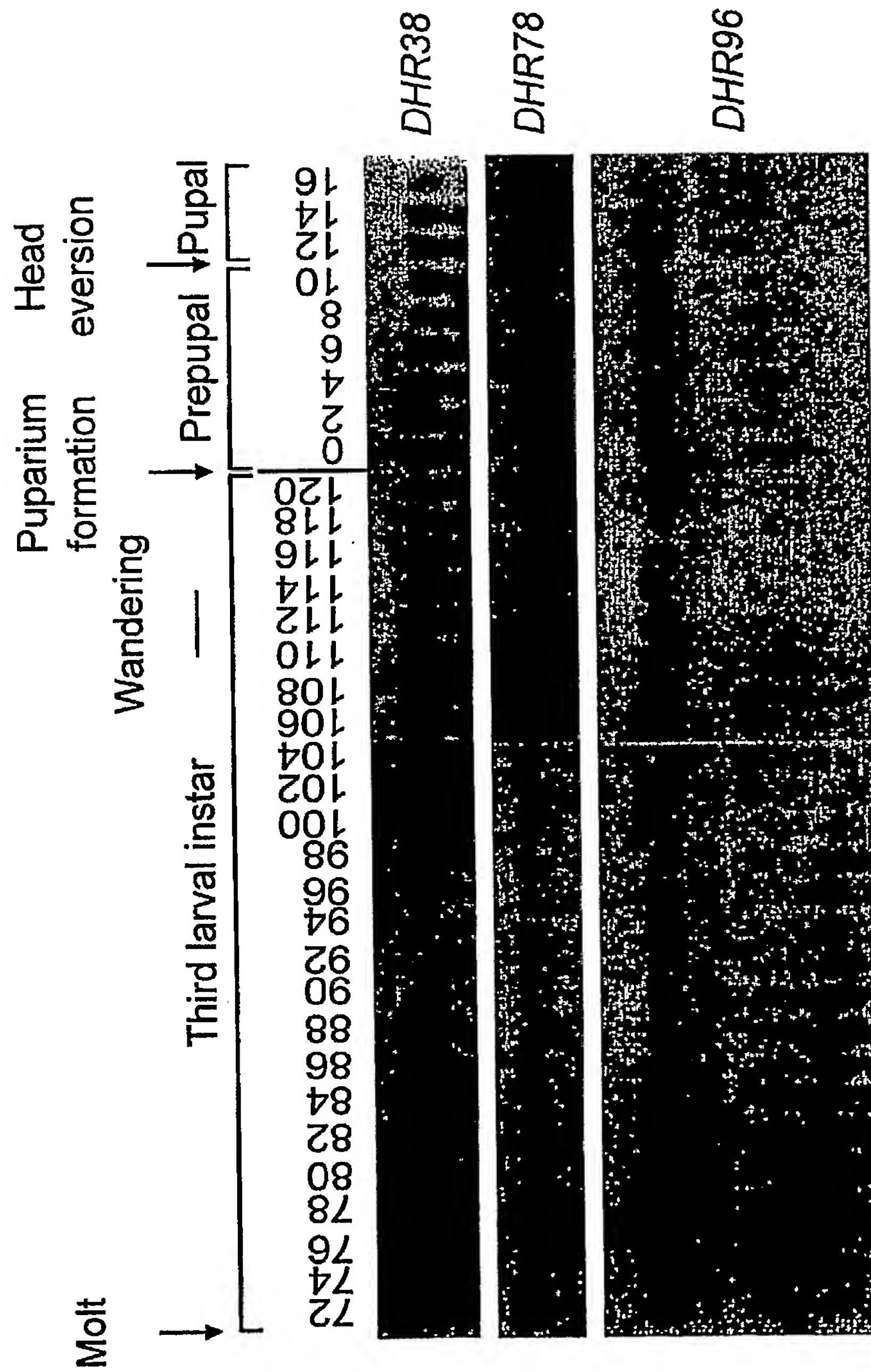


FIG. 6B

THR96	CAVCGDKALGYNFNAVTCESCKAFFRRNALAKKQ	FTCPFNQNCDIITVTRRFQCQCRLRKCLDIGM	Identity
TVDR	CGVCGDRATGCFHFNAMTCEGCKGFFRRSMKRKAL	FTCPFNQNCRIITKDNRRHCQA	64%
TCR	CLVCGDRASGYHYNALTCEGCKGFFRRSVTKSAV	YCCKFGRACEMDMYMRKCCQE	52%
ITZ-F1	CPVCGDKVSGYHYGLLTCEGCKGFFKRTVQNKKV	YTCVAERSCHIDKTQKRKCPY	49%
ACOUPTF	CVVCGDKSSGKHYGQFTCEGCKSFFKRSVRRNL	YTCRANRNCPIIDQHHRNQCCY	49%
TRARα	CFVGNDKSSGYHYGVSSCEGCKGFFRRSIQKNMV	YTCHRDKNCIINKVTRNRCCY	49%
IT3Rα	CVVCGDKATGYHYRCITCEGCKGFFRRTIQKNLHPTYS	CKYDSCCVIDKII TRNQCCQL	49%

FIG.6C



**FIG. 7**

9/16

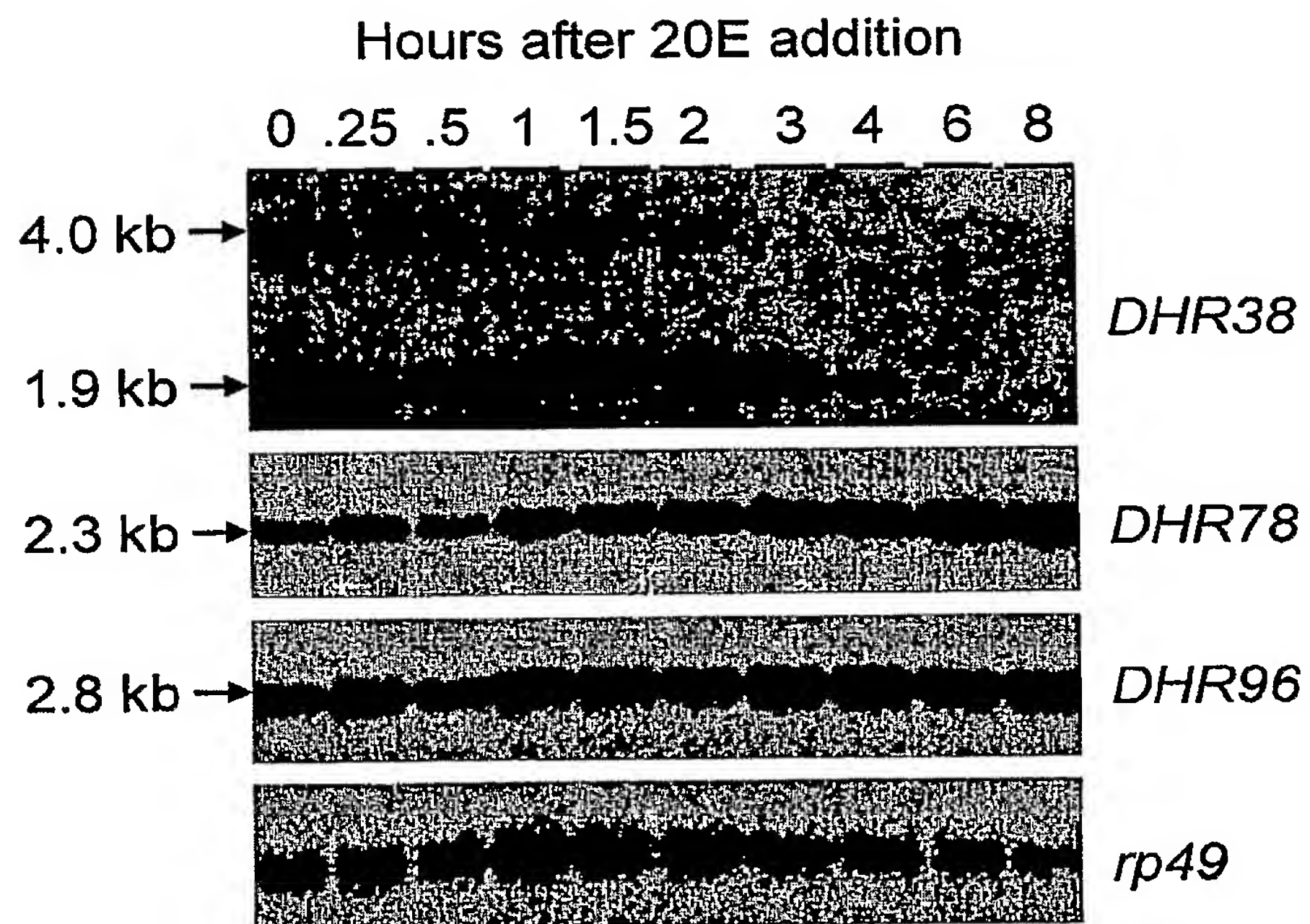


FIG.8

10/16

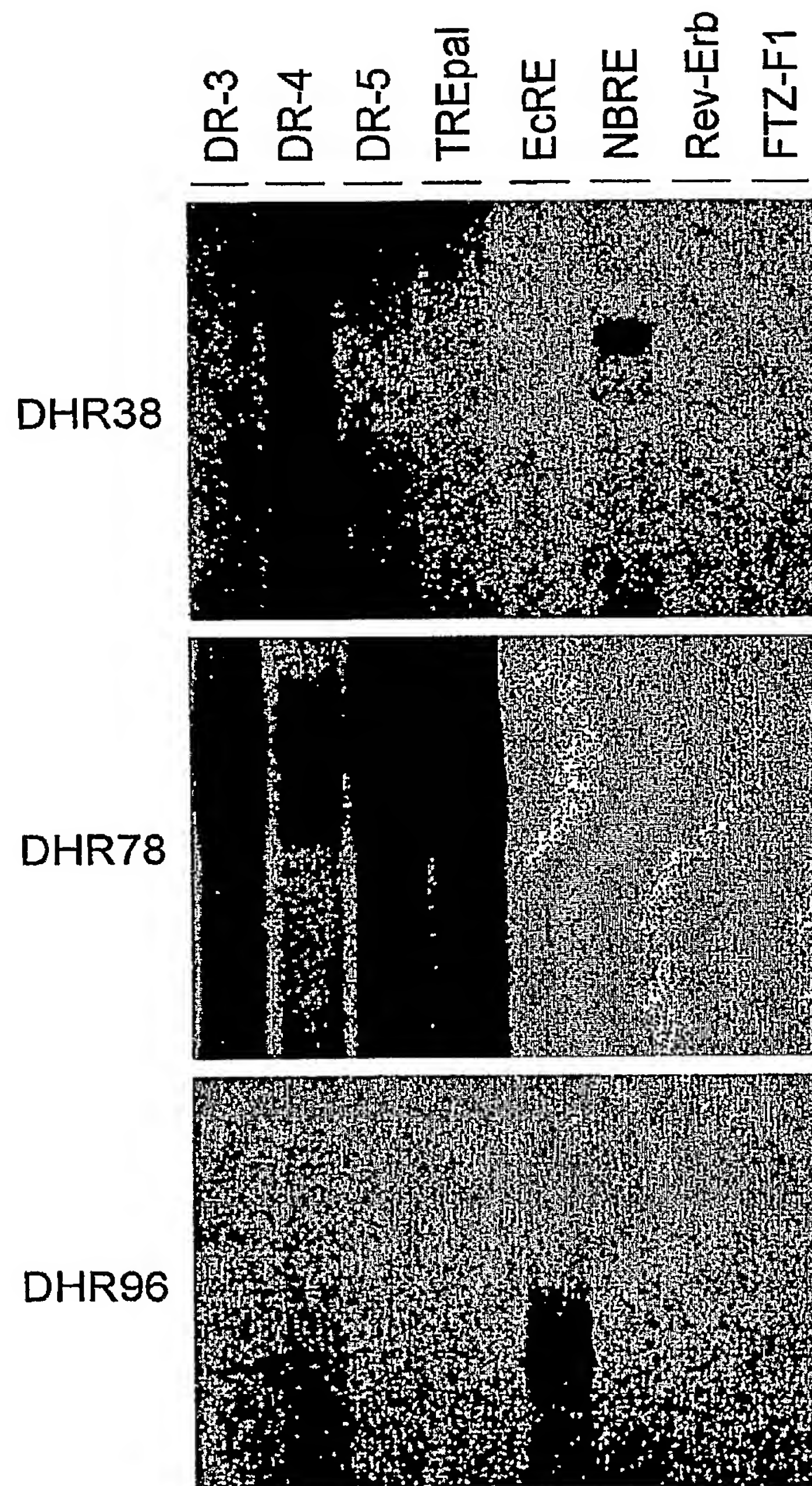


FIG.9

11/16

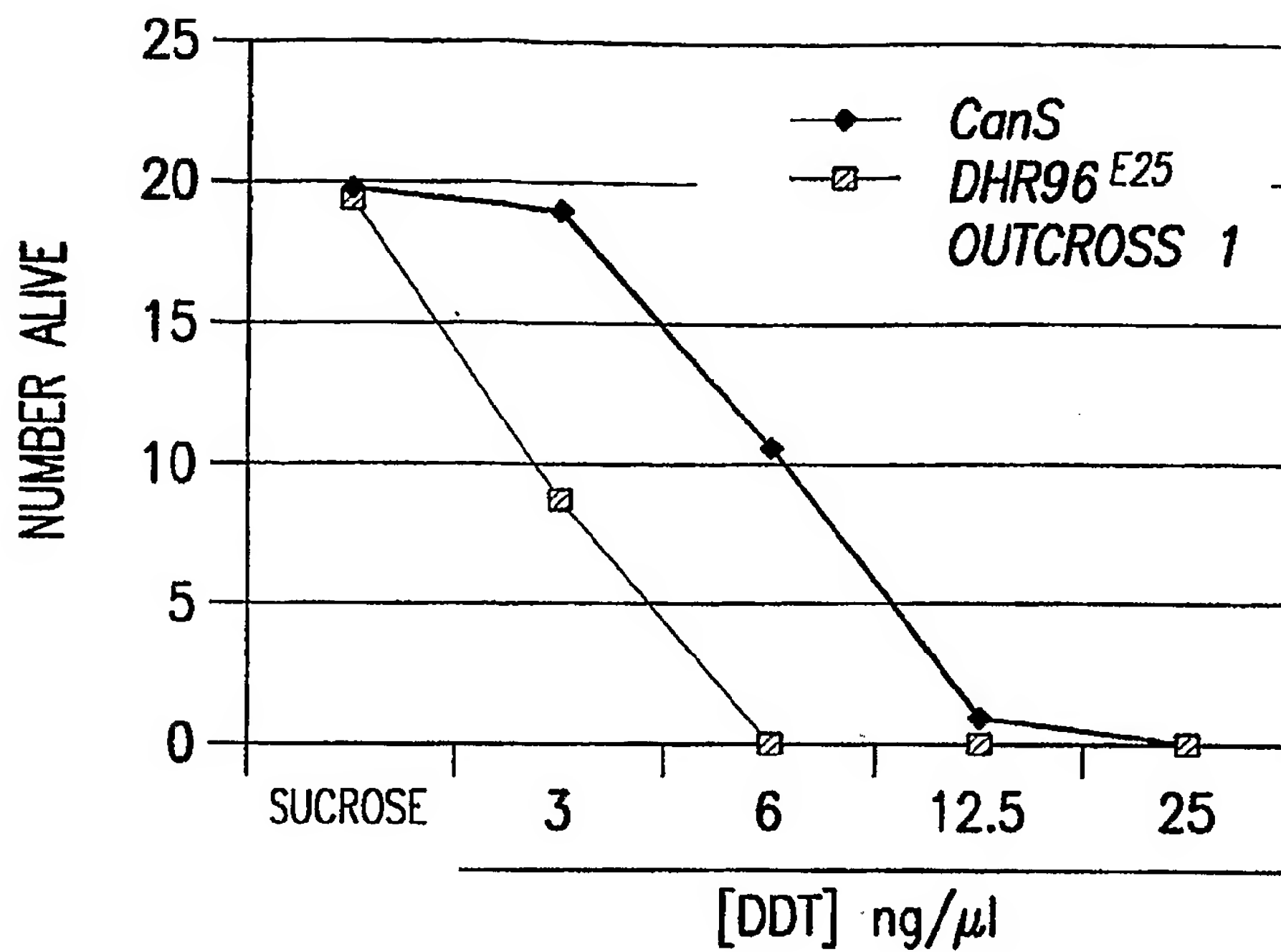


FIG. 10A

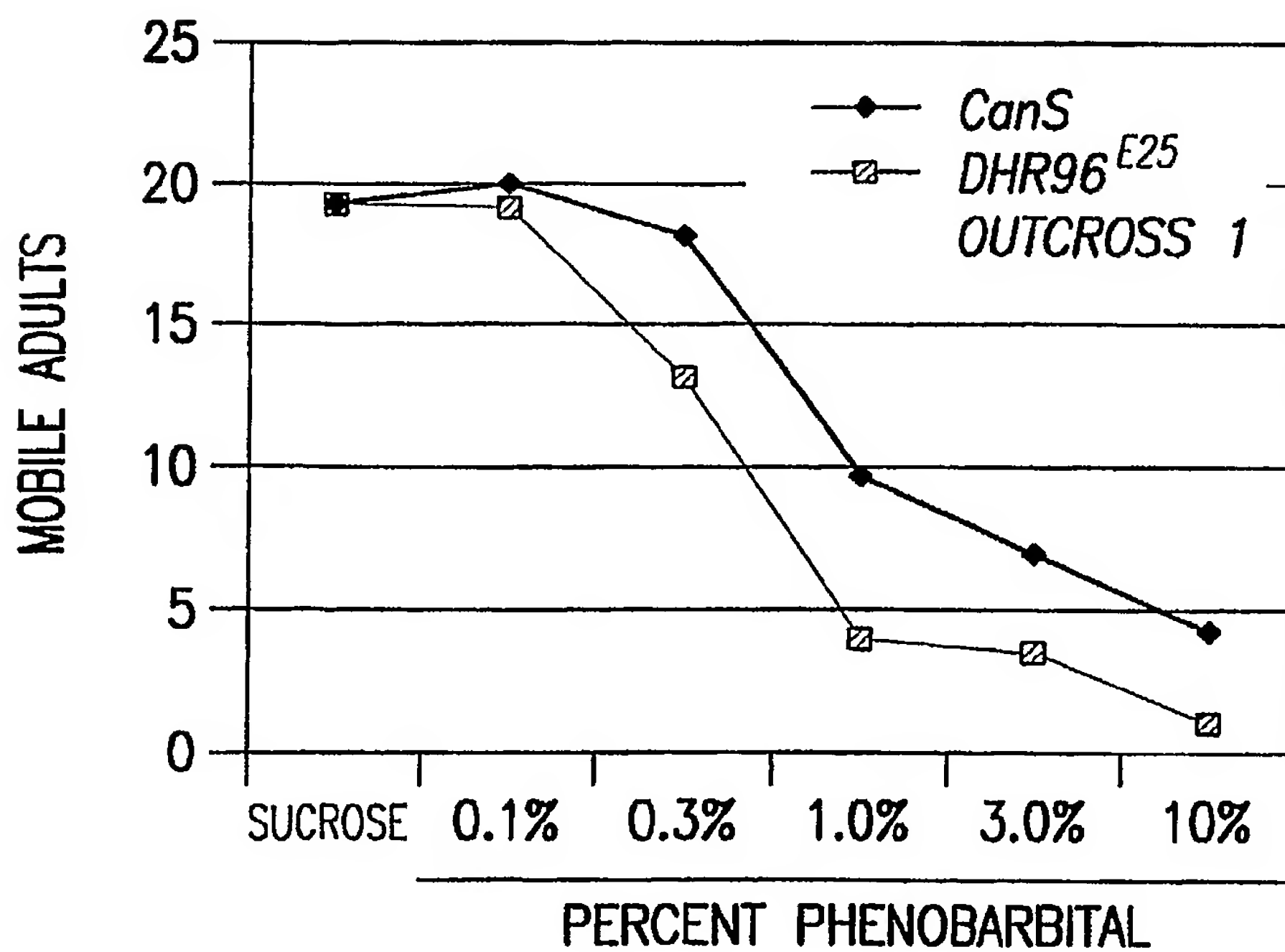


FIG. 10B

SUBSTITUTE SHEET (RULE 26)

12/16

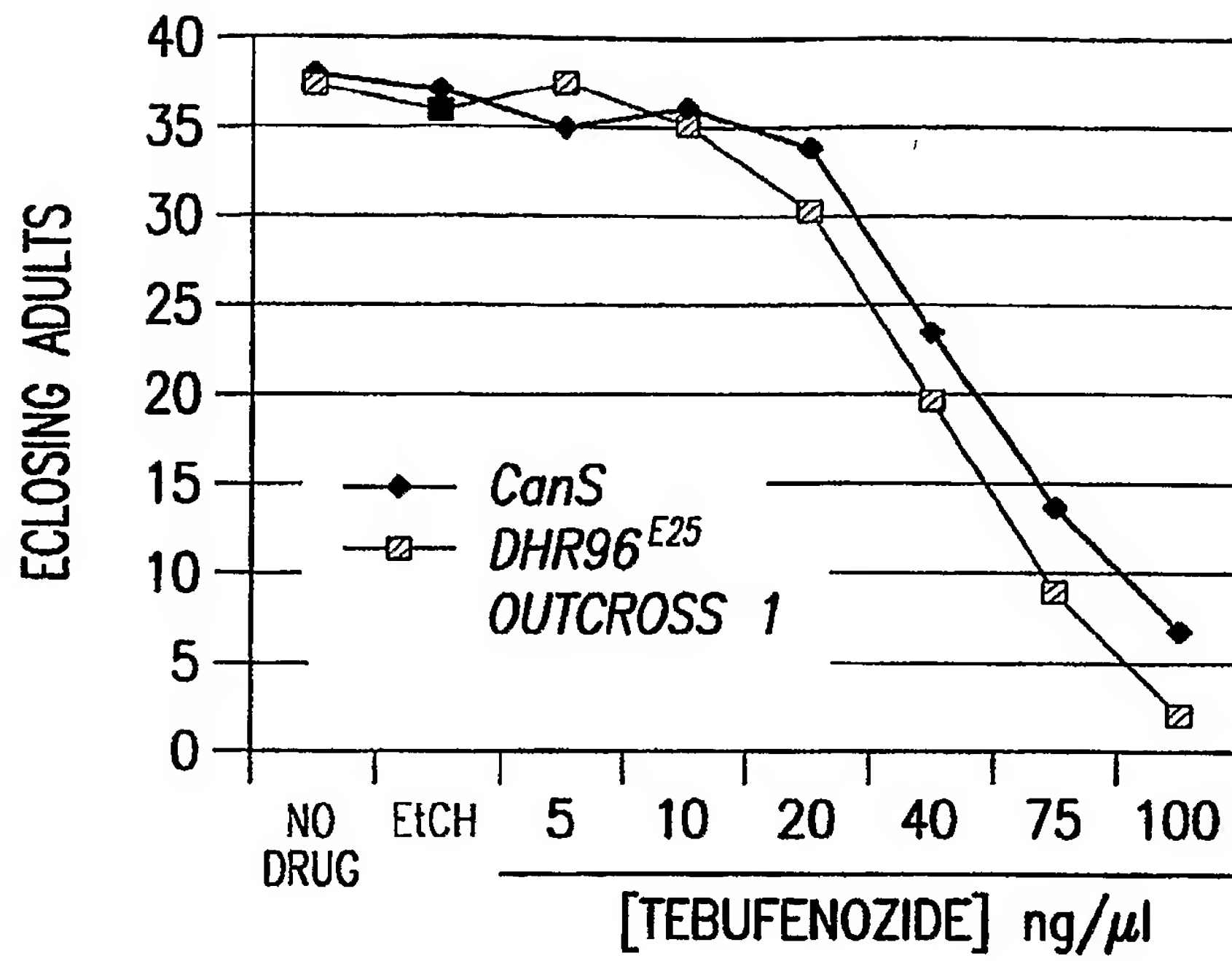


FIG.10C

description	WMPHS	96WPHS	fold change
3:FBgn0002564/sym=Lsp 1 gamma/name=Larva1 serum protein 1 gamma-s	11164.31	340.12	-32.82
3:FBgn0002531/sym=Lcp1/name=Larva1 cuticle protein 1/func=structural protein c	3369.11	513.4	-6.56
3:FBgn0019643/sym=Dat/name=Dopamine N acetyltransferase/prod=arylalkylamine N-acetyltransferase/func=	8868.78	1353.56	-6.55
3:FBgn0033596/sym=CG7738/name=/prod=/func=/map=47E1-47E1/transc=CT23557/len=405/GB:AE003:	4306.66	925.55	-4.65
3:FBgn0030484/sym=CG1681/name=/prod=glutathione transferase-like/func=enzyme/map=11F1-11F1/trans	946.93	221.04	-4.28
3:FBgn0040609/sym=CG3348/name=/prod=/func=map=97F6-97F6/transc=CT11231/len=516/GB:AE003:	5850.94	1473.79	-3.97
3:FBgn0039778/sym=CG18030/name=/prod=trypsin-like/func=endopeptidase/map=99F8-99F8/1transc=CT4:	1379.59	363.05	-3.8
3:FBgn0034294/sym=CG5765/name=/prod=/func=/map=55B1-55B1/transc=CT18116/len=1528/GB:AE00:	4602.49	1375.71	-3.35
3:FBgn0039296/sym=CG10420/name=/prod=/func=/map=96C3-96C3/transc=T29244/len=1794/GB:AE0	1271.55	405.9	-3.13
3:FBgn0002533/sym=Lcp2/name=Larva1 cuticle protein 2/prod=Larva1 cuticle protein 2/func=structural protein c	2050.81	723.13	-2.84
3:FBgn0015039/sym=Cyp9b2/name=Cytochrome P45-9b2/prod=cytochrome P450 CYP9B2/func=cytochr	2526.49	947.68	-2.67
3:FBgn0033978/sym=Cyp6a23/name=/prod=cytochrome P450, CYP6A23/func=cytochrome P45/map=51C	647.32	243.36	-2.66
3:FBgn0033595/sym=CG18337/name=/prod/func=/map=47E1-47E1/transc=CT41641/len=540 /GB:AE00:	538.81	202.16	-2.66
3:FBgn0036656/sym=CG 13026/name= /prod= /func= /map=73B5-73B5/transc=CT32244/len=405/GB:AE00	2244.66	892.89	-2.51
3:FBgn0000153/sym=b/name=b/ack/prod=glutamate decarboxylase 2/func=glutamate decarboxylase : EC:4.1	7070.57	2889.86	-2.45
3:FBgn0039798/sym=CG11313/name=/prod=monophenol monooxygenase activator-like func=endopeptidase	1080.32	451.23	-2.39
3:FBgn0032606/sym=CG17932/name=/prod=UDP-glucuronosyltransferase/func=enzyme/map=36A9-36A9A	1214.19	515	-2.36
3:FBgn0000594/sym=Est-P/name=Esterase P/prod=carboxylesterase/func=carboxylesterase: EC:3.1.1.1/map	475.26	205.79	-2.31
3:FBgn0039239/sym=CG13641/name=/prod=/func=/map=96B4-96B4/transc=CT33035/len=429/GB:AE00:	1531.71	717.27	-2.14
B:FBgn0014849/sym=Eig71Ei/name=/prod=/func=/map=71E4-71E4/transc=CT22591/len=3901GB:AE0035	1107.37	523.8	-2.11
B:FBgn0034341/sym=CG17531/name=GslE7/prod=glutathione transferase/func=enzyme /map=55C9-55C9/f	3476.77	1645.19	-2.11
B:FBgn0033830/sym=CG10814/name=/prod=/func=vitamin biosynthesis/map=50A6-50A6/transc=CT30312	363.72	174.9	-2.08

FIG.11A

Accession	Gene Name	Protein Name	Protein Description	Length	Mass (kDa)	pI	Charge
FBgn0034010	serine protease 1	Ser990a	serine protease 1 /prod=serine endopeptidase/func=serine carboxypept	1622.11	787.24	-2.06	-2.06
FBgn0033356	serine protease 2	Ser990b	serine protease 2 /prod=serine endopeptidase/func=serine-type endope	15597.45	7724.85	-2.03	-2.03
FBgn0001256	ecdysone-inducible gene L1	ImpLt	ecdysone-inducible secreted membrane	1414.79	709.8	-1.99	-1.99
FBgn0003046	pupal cuticle protein	Pcp	pupal cuticle protein of pupal cuticle (Drosophi	2111.66	1080.62	-1.95	-1.95
FBgn0003357	serine protease 2	Ser990b	serine protease 2 /prod=serine endopeptidase/func=serine-type endope	15454.1	1955.78	-1.94	-1.94
FBgn0036659	beta-glucosidase-like	CG9701	beta-glucosidase-like /func=ion channel /map=73B5-7385 /transc	1162.58	598.76	-1.94	-1.94
FBgn0033821	serine protease 1	CG10799	serine protease 1 /prod=serine endopeptidase/func=serine-type endope	822.24	425.81	-1.93	-1.93
FBgn0036024	serine protease 1	CG18180	serine protease 1 /prod=serine endopeptidase/func=serine-type endope	4368.44	2288.8	-1.91	-1.91
FBgn0013307	ornithine decarboxylase 1	Odc 1	ornithine decarboxylase 1 /prod=ornithine decarboxylase func=ornithine decr	1007.84	530.1	-1.9	-1.9
FBgn0029898	permease-like	CG14439	permease-like /func=transporter/map=6C11-6C11/transc=CT3	2393.73	1275.57	-1.88	-1.88
FBgn0031653	serine protease 1	CG8871	serine protease 1 /prod=serine endopeptidase/func=serine-type endope	959.63	512.75	-1.87	-1.87
FBgn0040565	serine protease 1	CG7606	serine protease 1 /prod=serine endopeptidase/func=serine-type endope	8325.78	4504.75	-1.85	-1.85
FBgn0031251	motor	CG4213	motor /func=motor/map=21C2-21C2/transc=CT13888/len=3696/GB:	146.23	19.37	-1.84	-1.84
FBgn0033197	serine protease-like	CG17984	serine protease-like /func=serine protease-like /map=43E6-43E6/transc=CT40154/len=1417/GB:AE0	1026.78	559.64	-1.83	-1.83
FBgn0039777	serine protease-like	CG2229	serine protease-like /func=serine protease-like /map=99F7-99F8/trans	9394.5	5150.37	-1.82	-1.82
FBgn0032889	glycerate dehydrogenase-like	CG9331	glycerate dehydrogenase-like /func=enzyme/map=38E9-38E9/trans	3023.89	1676.12	-1.8	-1.8

**FIG. 11B**

15/16

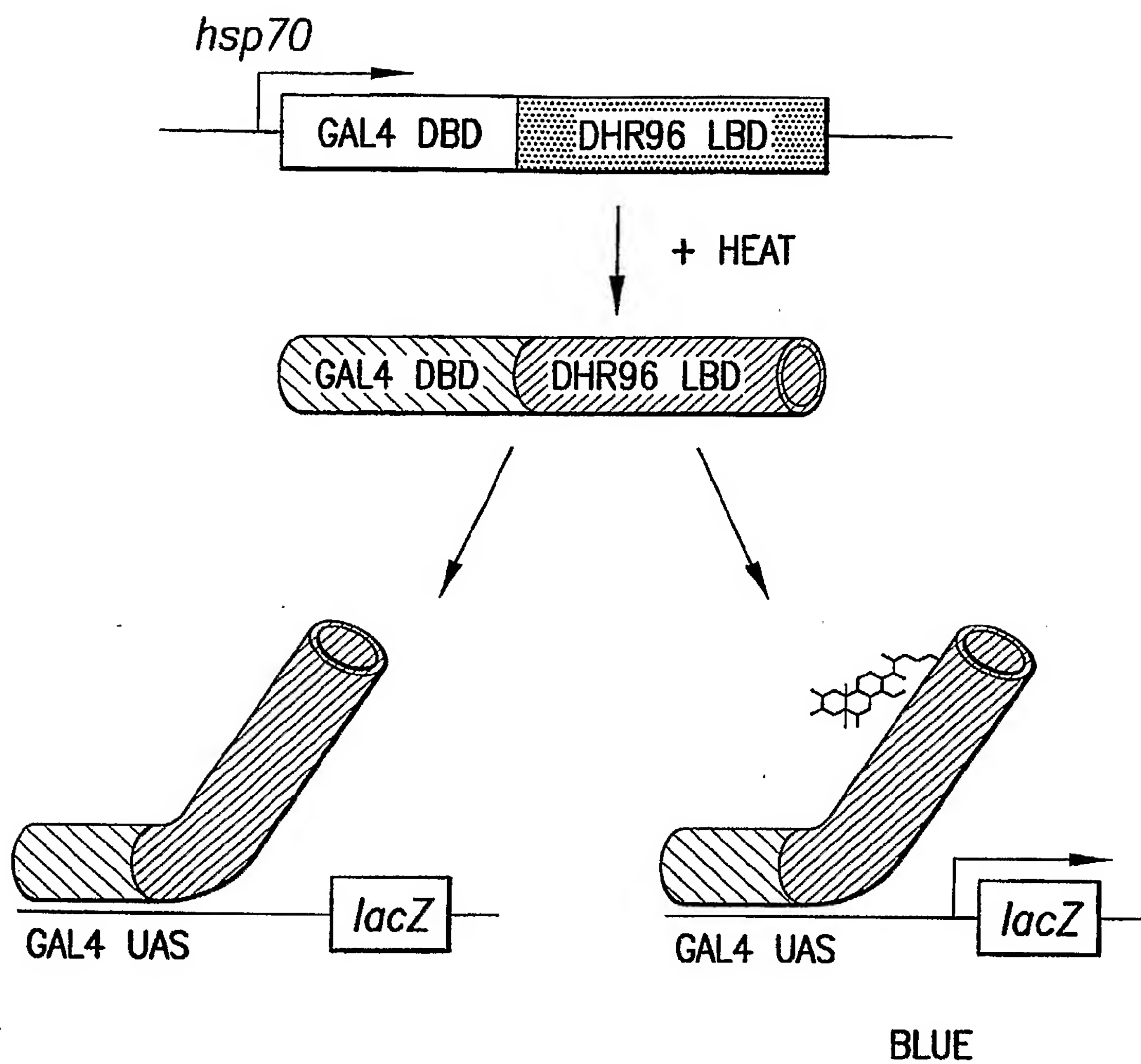


FIG. 12

16/16

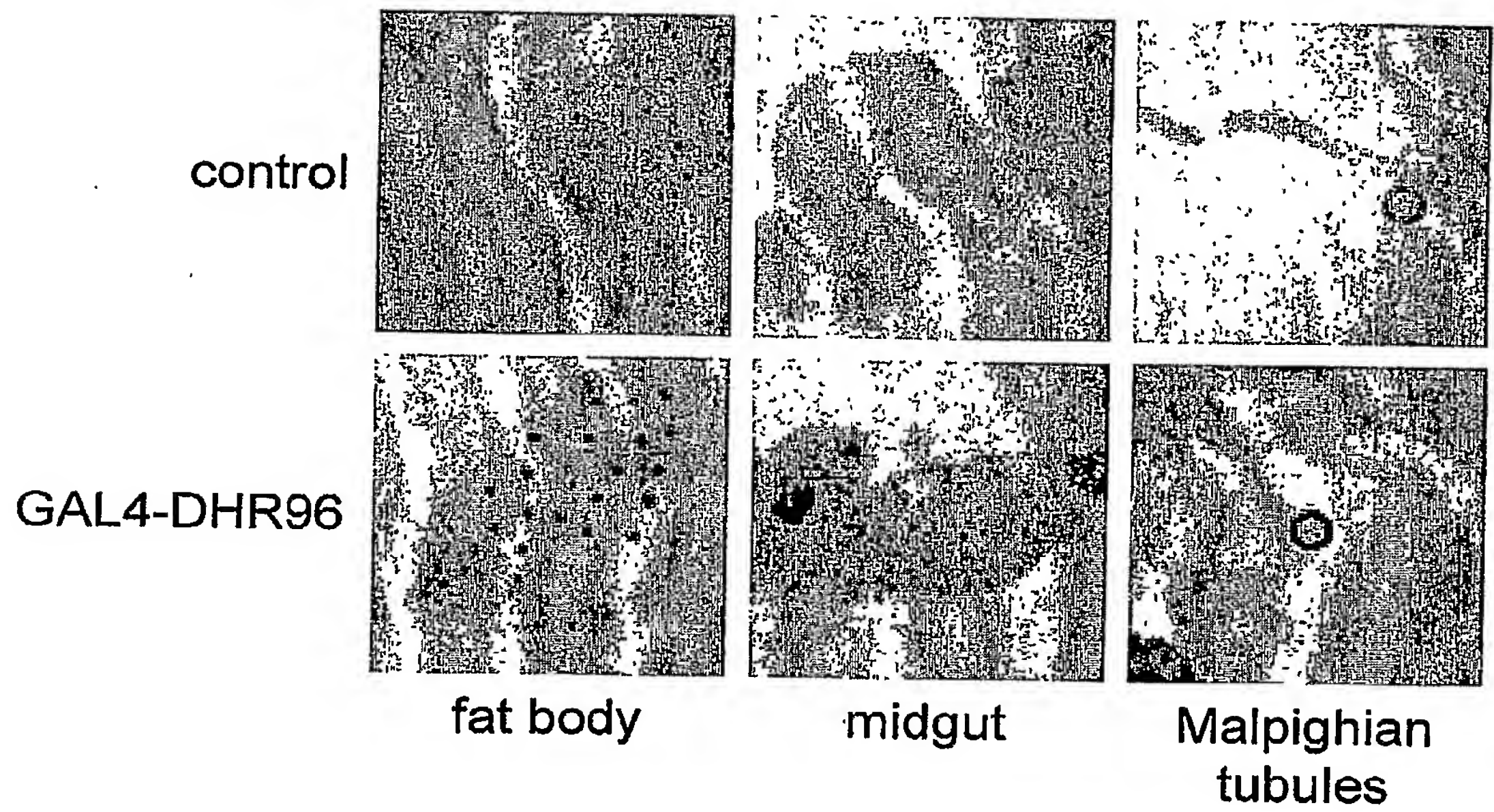


FIG.13

## SEQUENCE LISTING

&lt;110&gt; University of Utah Research Foundation

<120> COMPOSITIONS AND METHODS FOR MODULATING  
DHR96

&lt;130&gt; 21101.0053P1

&lt;140&gt; Unassigned

&lt;141&gt; 2005-01-13

&lt;150&gt; 60/536,337

&lt;151&gt; 2004-01-13

&lt;160&gt; 60

&lt;170&gt; FastSEQ for Windows Version 4.0

&lt;210&gt; 1

&lt;211&gt; 1543

&lt;212&gt; PRT

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence; note =  
synthetic construct

&lt;400&gt; 1

Met	Thr	Leu	Ser	Arg	Gly	Pro	Tyr	Ser	Glu	Leu	Asp	Lys	Met	Ser	Leu	1	5	10	15
Phe	Gln	Asp	Leu	Lys	Leu	Lys	Arg	Arg	Lys	Ile	Asp	Ser	Arg	Cys	Ser	20	25	30	
Ser	Asp	Gly	Glu	Ser	Ile	Ala	Asp	Thr	Ser	Thr	Ser	Ser	Pro	Asp	Leu	35	40	45	
Leu	Ala	Pro	Met	Ser	Pro	Lys	Leu	Cys	Asp	Ser	Gly	Ser	Ala	Gly	Ala	50	55	60	
Ser	Leu	Gly	Ala	Ser	Leu	Pro	Leu	Pro	Leu	Ala	Leu	Pro	Leu	Pro	Met	65	70	75	80
Ala	Leu	Pro	Leu	Pro	Met	Ser	Leu	Pro	Leu	Pro	Leu	Thr	Ala	Ala	Ser	85	90	95	
Ser	Ala	Val	Thr	Val	Ser	Leu	Ala	Ala	Val	Val	Ala	Ala	Val	Ala	Glu	100	105	110	
Thr	Gly	Gly	Ala	Gly	Ala	Gly	Gly	Ala	Gly	Thr	Ala	Val	Thr	Ala	Ser	115	120	125	
Gly	Ala	Gly	Pro	Cys	Val	Ser	Thr	Ser	Ser	Thr	Thr	Ala	Ala	Ala	Ala	130	135	140	
Thr	Ser	Ser	Thr	Ser	Ser	Leu	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	145	150	155	160
Thr	Ser	Ser	Ser	Thr	Ser	Ser	Ala	Ser	Pro	Thr	Ala	Gly	Ala	Ser	Ser	165	170	175	
Thr	Ala	Thr	Cys	Pro	Ala	Ser	Ser	Ser	Ser	Ser	Ser	Gly	Asn	Gly	Ser	180	185	190	
Gly	Gly	Lys	Ser	Gly	Ser	Ile	Lys	Gln	Glu	His	Thr	Glu	Ile	His	Ser	195	200	205	
Ser	Ser	Ser	Ala	Ile	Ser	Ala	Ala	Ala	Ala	Ser	Thr	Val	Met	Ser	Pro	210	215	220	
Pro	Pro	Ala	Glu	Ala	Thr	Arg	Ser	Ser	Pro	Ala	Thr	Pro	Glu	Gly	Gly	225	230	235	240

Gly	Pro	Ala	Gly	Asp	Gly	Ser	Gly	Ala	Thr	Gly	Gly	Gly	Asn	Thr	Ser			
				245					250					255				
Gly	Gly	Ser	Thr	Ala	Gly	Val	Ala	Ile	Asn	Glu	His	Gln	Asn	Asn	Gly			
			260					265					270					
Asn	Gly	Ser	Gly	Gly	Ser	Ser	Arg	Ala	Ser	Pro	Asp	Ser	Leu	Glu	Glu			
		275					280					285						
Lys	Pro	Ser	Thr	Thr	Thr	Thr	Thr	Gly	Arg	Pro	Thr	Leu	Thr	Pro	Thr			
	290					295					300							
Asn	Gly	Val	Leu	Ser	Ser	Ala	Ser	Ala	Gly	Thr	Gly	Ile	Ser	Thr	Gly			
305					310					315					320			
Ser	Ser	Ala	Lys	Leu	Ser	Glu	Ala	Gly	Met	Ser	Val	Ile	Arg	Ser	Val			
			325					330						335				
Lys	Glu	Glu	Arg	Leu	Leu	Asn	Val	Ser	Ser	Lys	Met	Leu	Val	Phe	His			
			340					345					350					
Gln	Gln	Arg	Glu	Gln	Glu	Thr	Lys	Ala	Val	Ala	Ala	Ala	Ala	Ala	Ala			
		355					360					365						
Ala	Ala	Ala	Gly	His	Val	Thr	Val	Leu	Val	Thr	Pro	Ser	Arg	Ile	Lys			
		370				375					380							
Ser	Glu	Pro	Pro	Pro	Pro	Ala	Ser	Pro	Ser	Ser	Thr	Ser	Ser	Thr	Gln			
385					390					395					400			
Arg	Glu	Arg	Glu	Arg	Glu	Arg	Asp	Arg	Glu	Arg	Asp	Arg	Glu	Arg	Glu			
			405					410						415				
Arg	Glu	Arg	Asp	Arg	Asp	Arg	Glu	Arg	Glu	Arg	Glu	Gln	Ser	Ile	Ser			
			420					425					430					
Ser	Ser	Gln	Gln	His	Leu	Ser	Arg	Val	Ser	Ala	Ser	Pro	Pro	Thr	Gln			
		435					440					445						
Leu	Ser	His	Gly	Ser	Leu	Gly	Pro	Asn	Ile	Val	Gln	Thr	His	His	Leu			
	450					455					460							
His	Gln	Gln	Leu	Thr	Gln	Pro	Leu	Thr	Leu	Arg	Lys	Ser	Ser	Pro	Pro			
465					470					475					480			
Thr	Glu	His	Leu	Leu	Ser	Gln	Ser	Met	Gln	His	Leu	Thr	Gln	Gln	Gln			
			485					490						495				
Ala	Ile	His	Leu	His	His	Leu	Leu	Gly	Gln	Gln	Gln	Gln	Gln	Gln	Gln			
			500					505					510					
Ala	Ser	His	Pro	Gln	Gln	Gln	Gln	Gln	Gln	Gln	His	Ser	Pro	His	Ser			
		515					520					525						
Leu	Val	Arg	Val	Lys	Lys	Glu	Pro	Asn	Val	Gly	Gln	Arg	His	Leu	Ser			
	530					535					540							
Pro	His	His	Gln	Gln	Gln	Ser	Pro	Leu	Leu	Gln	His	His	Gln	Gln	Gln			
545					550					555					560			
Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	His	Leu	His	Gln	Gln	Gln	Gln			
			565						570					575				
Gln	Gln	Gln	His	His	Gln	Gln	Gln	Pro	Gln	Ala	Leu	Ala	Leu	Met	His			
			580					585					590					
Pro	Ala	Ser	Leu	Ala	Leu	Arg	Asn	Ser	Asn	Arg	Asp	Ala	Ala	Ile	Leu			
		595					600					605						
Phe	Arg	Val	Lys	Ser	Glu	Val	His	Gln	Gln	Val	Ala	Ala	Gly	Leu	Pro			
	610					615					620							
His	Leu	Met	Gln	Ser	Ala	Gly	Gly	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala			
625					630					635					640			
Val	Ala	Ala	Gln	Arg	Met	Val	Cys	Phe	Ser	Asn	Ala	Arg	Ile	Asn	Gly			
			645					650						655				
Val	Lys	Pro	Glu	Val	Ile	Gly	Gly	Pro	Leu	Gly	Asn	Leu	Arg	Pro	Val			
			660					665					670					
Gly	Val	Gly	Gly	Gly	Asn	Gly	Ser	Gly	Ser	Val	Gln	Cys	Pro	Ser	Pro			
		675				680						685						
His	Pro	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Gln	Leu	Ser	Pro	Gln	Thr	Pro			
	690					695					700							
Ser	Gln	Thr	Pro	Pro	Arg	Gly	Thr	Pro	Thr	Val	Ile	Met	Gly	Glu	Ser			
705					710					715					720			

Cys	Gly	Val	Arg	Thr	Met	Val	Trp	Gly	Tyr	Glu	Pro	Pro	Pro	Pro	Ser	
				725					730						735	
Ala	Gly	Gln	Ser	His	Gly	Gln	His	Pro	Gln	Gln	Gln	Gln	Gln	Ser	Pro	
			740					745						750		
His	His	Gln	Pro	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Ser	Gln	
		755					760						765			
Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Ser	Leu	Gly	Gln	Gln	Gln	His	
	770					775					780					
Cys	Leu	Ser	Ser	Pro	Ser	Ala	Gly	Ser	Leu	Thr	Pro	Ser	Ser	Ser	Ser	
785					790					795					800	
Gly	Gly	Gly	Ser	Val	Ser	Gly	Gly	Gly	Val	Gly	Gly	Pro	Leu	Thr	Pro	
				805					810						815	
Ser	Ser	Val	Ala	Pro	Gln	Asn	Asn	Glu	Glu	Ala	Ala	Gln	Leu	Leu	Leu	
			820					825					830			
Ser	Leu	Gly	Gln	Thr	Arg	Ile	Gln	Asp	Met	Arg	Ser	Arg	Pro	His	Pro	
		835					840					845				
Phe	Arg	Thr	Pro	His	Ala	Leu	Asn	Met	Glu	Arg	Leu	Trp	Ala	Gly	Asp	
	850					855					860					
Tyr	Ser	Gln	Leu	Pro	Pro	Gly	Gln	Leu	Gln	Ala	Leu	Asn	Leu	Ser	Ala	
865					870					875					880	
Gln	Gln	Gln	Gln	Trp	Gly	Ser	Ser	Asn	Ser	Thr	Gly	Leu	Gly	Gly	Val	
				885					890						895	
Gly	Gly	Gly	Met	Gly	Gly	Arg	Asn	Leu	Glu	Ala	Pro	His	Glu	Pro	Thr	
			900					905					910			
Asp	Glu	Asp	Glu	Gln	Pro	Leu	Val	Cys	Met	Ile	Cys	Glu	Asp	Lys	Ala	
		915					920					925				
Thr	Gly	Leu	His	Tyr	Gly	Ile	Ile	Thr	Cys	Glu	Gly	Cys	Lys	Gly	Phe	
	930					935					940					
Phe	Lys	Arg	Thr	Val	Gln	Asn	Arg	Arg	Val	Tyr	Thr	Cys	Val	Ala	Asp	
945					950					955					960	
Gly	Thr	Cys	Glu	Ile	Thr	Lys	Ala	Gln	Arg	Asn	Arg	Cys	Gln	Tyr	Cys	
				965					970					975		
Arg	Phe	Lys	Lys	Cys	Ile	Glu	Gln	Gly	Met	Val	Leu	Gln	Ala	Val	Arg	
			980					985					990			
Glu	Asp	Arg	Met	Pro	Gly	Gly	Arg	Asn	Ser	Gly	Ala	Val	Tyr	Asn	Leu	
		995					1000					1005				
Tyr	Lys	Val	Lys	Tyr	Lys	Lys	His	Lys	Lys	Thr	Asn	Gln	Lys	Gln	Gln	
	1010					1015					1020					
Gln	Gln	Ala	Ala	Gln	Gln	Gln	Gln	Gln	Gln	Ala	Ala	Ala	Gln	Gln	Gln	
1025					1030					1035					1040	
His	Gln	Gln	Gln	Gln	Gln	His	Gln	Gln	His	Gln	Gln	His	Gln	Gln	Gln	
				1045					1050						1055	
Gln	Leu	His	Ser	Pro	Leu	His	His	His	His	His	Gln	Gly	His	Gln	Ser	
			1060					1065					1070			
His	His	Ala	Gln	Gln	Gln	His	His	Pro	Gln	Leu	Ser	Pro	His	His	Leu	
		1075					1080					1085				
Leu	Ser	Pro	Gln	Gln	Gln	Gln	Leu	Ala	Ala	Ala	Val	Ala	Ala	Ala	Ala	
	1090					1095					1100					
Gln	His	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Ala	
1105					1110					1115					1120	
Lys	Leu	Met	Gly	Gly	Val	Val	Asp	Met	Lys	Pro	Met	Phe	Leu	Gly	Pro	
			1125						1130					1135		
Ala	Leu	Lys	Pro	Glu	Leu	Leu	Gln	Ala	Pro	Pro	Met	His	Ser	Pro	Ala	
			1140					1145					1150			
Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Ala	Ser	
		1155					1160					1165				
Pro	His	Leu	Ser	Leu	Ser	Ser	Pro	His	Gln	Gln	Gln	Gln	Gln	Gln	Gln	
	1170					1175				1180						
Gly	Gln	His	Gln	Asn	His	His	Gln	Gln	Gln	Gly	Gly	Gly	Gly	Gly	Gly	
1185				1190						1195					1200	

Ala Gly Gly Gly Ala Gln Leu Pro Pro His Leu Val Asn Gly Thr Ile  
1205 1210 1215  
Leu Lys Thr Ala Leu Thr Asn Pro Ser Glu Ile Val His Leu Arg His  
1220 1225 1230  
Arg Leu Asp Ser Ala Val Ser Ser Lys Asp Arg Gln Ile Ser Tyr  
1235 1240 1245  
Glu His Ala Leu Gly Met Ile Gln Thr Leu Ile Asp Cys Asp Ala Met  
1250 1255 1260  
Glu Asp Ile Ala Thr Leu Pro His Phe Ser Glu Phe Leu Glu Asp Lys  
1265 1270 1275 1280  
Ser Glu Ile Ser Glu Lys Leu Cys Asn Ile Gly Asp Ser Ile Val His  
1285 1290 1295  
Lys Leu Val Ser Trp Thr Lys Lys Leu Pro Phe Tyr Leu Glu Ile Pro  
1300 1305 1310  
Val Glu Ile His Thr Lys Leu Leu Thr Asp Lys Trp His Glu Ile Leu  
1315 1320 1325  
Ile Leu Thr Thr Ala Ala Tyr Gln Ala Leu His Gly Lys Arg Arg Gly  
1330 1335 1340  
Glu Gly Gly Gly Ser Arg His Gly Ser Pro Ala Ser Thr Pro Leu Ser  
1345 1350 1355 1360  
Thr Pro Thr Gly Thr Pro Leu Ser Thr Pro Ile Pro Ser Pro Ala Gln  
1365 1370 1375  
Pro Leu His Lys Asp Asp Pro Glu Phe Val Ser Glu Val Asn Ser His  
1380 1385 1390  
Leu Ser Thr Leu Gln Thr Cys Leu Thr Thr Leu Met Gly Gln Pro Ile  
1395 1400 1405  
Ala Met Glu Gln Leu Lys Leu Asp Val Gly His Met Val Asp Lys Met  
1410 1415 1420  
Thr Gln Ile Thr Ile Met Phe Arg Arg Ile Lys Leu Lys Met Glu Glu  
1425 1430 1435 1440  
Tyr Val Cys Leu Lys Val Tyr Ile Leu Leu Asn Lys Gly Thr Trp Phe  
1445 1450 1455  
Asp Leu Gln Asn Pro Phe Ile Gln Cys Ser Cys Tyr Leu Leu Val Arg  
1460 1465 1470  
Phe Val Asn Pro Ala Glu Val Glu Leu Glu Ser Ile Gln Glu Arg Tyr  
1475 1480 1485  
Val Gln Val Leu Arg Ser Tyr Leu Gln Asn Ser Ser Pro Gln Asn Pro  
1490 1495 1500  
Gln Ala Arg Leu Ser Glu Leu Leu Ser His Ile Pro Glu Ile Gln Ala  
1505 1510 1515 1520  
Ala Ala Ser Leu Leu Leu Glu Ser Lys Met Phe Tyr Val Pro Phe Val  
1525 1530 1535  
Leu Asn Ser Ala Ser Ile Arg  
1540

&lt;210&gt; 2

&lt;211&gt; 4632

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence; note =  
synthetic construct

&lt;400&gt; 2

atgacactga	gccgtggccc	gtacagcgag	ctcgataaaa	tgagcctttt	tcaagacctc	60
aaactcaaac	ggcgcaaaat	cgattcgcg	tgcagcagtg	acggcgagtc	catagcggac	120
acgtccacct	cgtcgccgga	cctgctggcg	cccatgtcgc	cgaagctctg	cgacagcggc	180
tcggcggggg	cgtcgctggg	ggcatcgctg	cccctgccgc	tggccctgcc	cctgccaatg	240
gccctgccac	tgcccatgtc	gctgcccttg	cccctcacgg	cggcatcttc	ggcggtcacc	300
gtttcgctgg	cagcggtcgt	ggccgcggtg	gccgagacgg	gtggcgcggg	cgcgaggagga	360

gctgggacag	cagtaacagc	gtcgggagca	ggaccatgcg	tctccacgtc	gtctacgacg	420
gcagcggcag	ccacatcctc	gacctcctcg	ctctcgtcct	cctcctcttc	gtcatcctcc	480
acgtcctcca	gcacttcctc	cgcctcgccg	acagctggag	cctcctccac	ggccacctgc	540
cccgccagca	gcagcagcag	cagtggaaac	ggaagtgggg	gcaaaagtgg	tagcatcaag	600
caggagcaca	cggagataca	ctcgtcgagc	agtgcgattt	cggcggccgc	cgcctcaacg	660
gtgatgtcac	cgccgcccgc	tgaggcgacg	agatccagtc	cagccacgcc	cgagggaggc	720
ggaccagctg	gcgacggaag	tggagcaacg	ggaggcgga	acacgagcgg	cggatcaacg	780
gctggagtgg	ccattaatga	acaccaaaac	aatggcaatg	gcagcggcgg	gagcagtcga	840
gcctctcccc	attcgctgga	agagaagccc	tctaccacaa	cgaccacagg	tcgtccaacg	900
ctcacgcccc	cgaatggggg	gctgtcctcc	gcctcggcgg	gcacggggat	ttccacagga	960
agcagcgcca	agctgagcga	ggctgggtatg	agtgtgatac	ggtccgtgaa	ggaggagcgc	1020
ttgctcaacg	tatccagcaa	gatgctgggtg	ttccatcagc	agcgggagca	agagacccaa	1080
gcagtggcgg	ctgcagcagc	agcagcagcg	gcggggccatg	tgacggttct	agtgcgcca	1140
tcgcgcatca	aatcggagcc	accgcccgcg	gcttcaccct	cctctacatc	cagcacacaa	1200
agggaaaggg	aacgggaacg	cgatcgagag	agggatcgcg	aaagggaacg	cgagcgggac	1260
cgggaccggg	aacgggaacg	ggaacagtcc	atcagctcct	cgcagcagca	cctaagtcgg	1320
gtctccgcca	gtccaccac	tcagctgtcc	cacggcagcc	tgggacccaa	cattgtgcag	1380
acgcaccatc	ttcaccagca	actcacacag	ccgctgacgc	tgcgcaagag	cagccccgcc	1440
acagagcacc	tgctcagtca	gtccatgcaa	catctcacac	agcagcaggc	gatccacctg	1500
catcacctac	ttggccagca	gcagcagcag	cagcaggcgt	cgcattccca	gcagcaacag	1560
cagcagcaac	actcgcccc	ctccctgggtg	cgggtgaaaa	aggaaccgaa	tgttgggtcag	1620
cggcacttat	cgccgcacat	ccaacaacag	tcgccactcc	tgcagcacca	ccaacagcag	1680
cagcagcagc	aacaacaaca	gcaacagcat	ctgcacacgc	aacagcaaca	gcagcagcat	1740
caccagcagc	agccccaggc	actggccctg	atgcattccg	cttccctggc	gctaagggaac	1800
agcaatcggg	atgcggccat	tctgtttcgg	gtgaagagcg	aagtgcacca	gcaggtggcc	1860
gccgggctgc	cgcattctgat	gcagtccgct	ggtggggcag	cggccgcgcg	cgcagcagct	1920
gtggccgctc	agcgaatggg	atgcttcagc	aatgccagga	tcaatggcgt	taagccggag	1980
gtgattggag	gaccgctggg	caacctgcgg	cccgtgggcg	tcggtggcgg	aaacggaagt	2040
ggctccgtgc	agtgcctcct	gccgcaccca	tcctcctcgt	cgtcatcctc	gcagctgtcg	2100
ccgcagacgc	cctcccagac	gccgcccoga	ggcacgcccc	ccgtcataat	gggcgagagc	2160
tgcggggtgc	gcaccatggg	ctgggggtac	gagcctccgc	caacctcggc	gggccagtcc	2220
cacggccagc	accgcgaaca	gcaacagcag	tcgccccacc	accagccgca	acaacaacag	2280
cagcagcaac	aacagcagtc	gcagcagcaa	cagcaacagc	agcagcaaca	gtcgctgggc	2340
cagcagcagc	actgcctctc	ctcgccgtcg	gcgggatcgc	tgacgccttc	ctcttcgtcc	2400
ggcgggtggt	cggatatctg	cggcggagtg	ggcggaccac	tcacaccctc	ctcgggtggcg	2460
ccgcagaata	acgaggaggc	cgcaccaactc	ctgctctccc	tgggacagac	acgcatccag	2520
gacatgagat	cacggccaca	ccccttcgcg	acaccgcacg	cccttaatat	ggagcggctg	2580
tgggcgggag	actactcgca	attgccgccc	ggccagctgc	aggctctgaa	tctcagtgcc	2640
caacagcagc	agtggggcag	cagcaactcc	acgggtcttg	gtggcgtagg	cggcggcatg	2700
ggcggacgca	acctggaggc	gccgcacgag	ccgaccgacg	aggacgaaca	gccgctcgtt	2760
tgcatgatct	gcgaggacaa	ggccaccggc	ctgcactacg	gcattcatcac	ctgcgagggg	2820
tgcaagggtc	tcttcaagcg	gacgggtgcag	aaccgacgag	tctacacctg	cgtggcggac	2880
ggcacctgcg	agataaccaa	agcacagcgc	aaccgttgtc	agtattgtcg	atttaagaag	2940
tgcatcgagc	agggcatggg	gctgcaagcc	gttcgcgagg	atcgcatgcc	gggcggtcgc	3000
aacagtggcg	ccgtctacaa	tttgtacaa	gtgaagtaca	agaagcaca	gaagaccaat	3060
cagaagcagc	agcagcaggc	cgcaccagcag	cagcagcagc	aggcggcggc	gcagcagcag	3120
caccagcaac	agcagcagca	tcaacagcac	cagcaacatc	agcaacagca	ggtgcactcg	3180
ccgctccacc	atcaccacca	ccaggggccac	cagtcgcacc	acgcgcagca	gcagcaccac	3240
ccacagctgt	cgcgcacca	cctgctgtcg	ccgcagcagc	agcaacttgc	cgcgcgggtg	3300
gcagcagctg	cgcagcacca	acagcaacag	caacaacagc	agcaacagca	gcagcaggcc	3360
aagctgatgg	gcggcgtggg	ggacatgaag	cccatgttcc	tcggccccgc	tttgaagccg	3420
gagttgctgc	aagcaccccc	catgcacagt	ccggcccagc	aacaacaaca	gcagcagcag	3480
cagcagcagc	aacagcaggc	ctcgccgcac	ctctcgctta	gctcacccga	ccagcagcag	3540
cagcagcagc	agggacagca	ccaaaaccac	caccagcaac	aaggtggggg	tggcggagga	3600
gctggtggag	gagctcaact	gccgcgcgac	ctggtgaacg	gaacgatact	gaagacggcc	3660
ctaaccaatc	ccagcgagat	tgtacatctg	cgcaccgcgc	tcgactcggc	ggtcagttcg	3720
tccaaggacc	gacagatctc	gtacgagcac	gccttaggca	tgatccagac	actgatcgac	3780
tgcgacgcga	tggaggacat	agccacactg	ccgcacttca	gcgagttcct	tgaggacaag	3840
tcggagatta	gcgagaaact	gtgcaacatc	ggcgattcca	tagtccacaa	gctggtgtcg	3900
tggacaaaaa	agttgccctt	ctacctggag	atcccggtgg	agatacatat	caaactactg	3960
acggacaagt	ggcacgagat	ccttatcctg	accacggccg	cctaccaggc	ggtgcatggc	4020

```

aagcggcgtg gcgagggagg aggcagcagg catggttcgc cggcgtcaac gccactgagc 4080
acgcccactg gtacgccgtt gagcacaccg ataccctcgc ccgcccagcc actgcacaag 4140
gacgacccgg agtttgtcag cgaggtgaac tcgcacctga gcacactgca aacctgcttg 4200
accacgctaa tgggccagcc gatagcgatg gagcagctga agctggacgt cgggcacatg 4260
gtggacaaga tgaccagat caccatcatg ttccggcgaa tcaagctcaa gatggaggag 4320
tacgtctgcc tgaaggttta catactgcta aacaaaggta cgtggttcga tttgcaaaac 4380
ccattcatac agtgctcatg ttaccttctc gttcgttttg taaatccagc agaagtggaa 4440
ctggagagca tccaggagcg gtacgtccag gtgctgcgct cctacctgca aaactcctcg 4500
ccgcagaatc cgcaggcgag gctcagtga ctgctctccc acataccaga gatccaggct 4560
gcggttagcc tgctgctcga gagcaagatg ttctatgtgc ccttcgtgct caactcggcg 4620
agcataaggt ag 4632

```

&lt;210&gt; 3

&lt;211&gt; 803

&lt;212&gt; PRT

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence; note =  
synthetic construct

&lt;400&gt; 3

```

Met Leu Leu Glu Met Asp Gln Gln Gln Ala Thr Val Gln Phe Ile Ser
 1          5          10          15
Ser Leu Asn Ile Ser Pro Phe Ser Met Gln Leu Glu Gln Gln Gln
          20          25          30
Pro Ser Ser Pro Ala Leu Ala Ala Gly Gly Asn Ser Ser Asn Asn Ala
          35          40          45
Ala Ser Gly Ser Asn Asn Asn Ser Ala Ser Gly Asn Asn Thr Ser Ser
          50          55          60
Ser Ser Asn Asn Asn Asn Asn Asn Asn Asn Asp Asn Asp Ala His Val
65          70          75          80
Leu Thr Lys Phe Glu His Glu Tyr Asn Ala Tyr Thr Leu Gln Leu Ala
          85          90          95
Gly Gly Gly Gly Ser Gly Ser Gly Asn Gln Gln His His Ser Asn His
          100          105          110
Ser Asn His Gly Asn His His Gln Gln Gln Gln Gln Gln Gln Gln
          115          120          125
Gln Gln Gln His Gln Gln Gln Gln Gln Glu His Tyr Gln Gln Gln Gln
          130          135          140
Gln Gln Asn Ile Ala Asn Asn Ala Asn Gln Phe Asn Ser Ser Ser Tyr
145          150          155          160
Ser Tyr Ile Tyr Asn Phe Asp Ser Gln Tyr Ile Phe Pro Thr Gly Tyr
          165          170          175
Gln Asp Thr Thr Ser Ser His Ser Gln Gln Ser Gly Gly Gly Gly Gly
          180          185          190
Gly Gly Gly Gly Asn Leu Leu Asn Gly Ser Ser Gly Gly Ser Ser Ala
          195          200          205
Gly Gly Gly Tyr Met Leu Leu Pro Gln Ala Ala Ser Ser Ser Gly Asn
          210          215          220
Asn Gly Asn Pro Asn Ala Gly His Met Ser Ser Gly Ser Val Gly Asn
225          230          235          240
Gly Ser Gly Gly Ala Gly Asn Gly Gly Ala Gly Gly Asn Ser Gly Pro
          245          250          255
Gly Asn Pro Met Gly Gly Thr Ser Ala Thr Pro Gly His Gly Gly Glu
          260          265          270
Val Ile Asp Phe Lys His Leu Phe Glu Glu Leu Cys Pro Val Cys Gly
          275          280          285
Asp Lys Val Ser Gly Tyr His Tyr Gly Leu Leu Thr Cys Glu Ser Cys
          290          295          300

```

Lys	Gly	Phe	Phe	Lys	Arg	Thr	Val	Gln	Asn	Lys	Lys	Val	Tyr	Thr	Cys
305					310					315					320
Val	Ala	Glu	Arg	Ser	Cys	His	Ile	Asp	Lys	Thr	Gln	Arg	Lys	Arg	Cys
				325					330						335
Pro	Tyr	Cys	Arg	Phe	Gln	Lys	Cys	Leu	Glu	Val	Gly	Met	Lys	Leu	Glu
				340					345						350
Ala	Val	Arg	Ala	Asp	Arg	Met	Arg	Gly	Gly	Arg	Asn	Lys	Phe	Gly	Pro
				355					360						365
Met	Tyr	Lys	Arg	Asp	Arg	Ala	Arg	Lys	Leu	Gln	Val	Met	Arg	Gln	Arg
				370					375						380
Gln	Leu	Ala	Leu	Gln	Ala	Leu	Arg	Asn	Ser	Met	Gly	Pro	Asp	Ile	Lys
385					390					395					400
Pro	Thr	Pro	Ile	Ser	Pro	Gly	Tyr	Gln	Gln	Ala	Tyr	Pro	Asn	Met	Asn
				405						410					415
Ile	Lys	Gln	Glu	Ile	Gln	Ile	Pro	Gln	Val	Ser	Ser	Leu	Thr	Gln	Ser
				420						425					430
Pro	Asp	Ser	Ser	Pro	Ser	Pro	Ile	Ala	Ile	Ala	Leu	Gly	Gln	Val	Asn
				435						440					445
Ala	Ser	Thr	Gly	Gly	Val	Ile	Ala	Thr	Pro	Met	Asn	Ala	Gly	Thr	Gly
				450											460
Gly	Ser	Gly	Gly	Gly	Gly	Leu	Asn	Gly	Pro	Ser	Ser	Val	Gly	Asn	Gly
465					470					475					480
Asn	Ser	Ser	Asn	Gly	Ser	Ser	Asn	Gly	Asn	Asn	Asn	Ser	Ser	Thr	Gly
				485						490					495
Asn	Gly	Thr	Ser	Gly	Gly	Gly	Gly	Gly	Asn	Asn	Ala	Gly	Gly	Gly	Gly
				500						505					510
Gly	Gly	Thr	Asn	Ser	Asn	Asp	Gly	Leu	His	Arg	Asn	Gly	Gly	Asn	Gly
				515											520
Asn	Ser	Ser	Cys	His	Glu	Ala	Gly	Ile	Gly	Ser	Leu	Gln	Asn	Thr	Ala
				530											535
Asp	Ser	Lys	Leu	Cys	Phe	Asp	Ser	Gly	Thr	His	Pro	Ser	Ser	Thr	Ala
545					550					555					560
Asp	Ala	Leu	Ile	Glu	Pro	Leu	Arg	Val	Ser	Pro	Met	Ile	Arg	Glu	Phe
				565											570
Val	Gln	Ser	Ile	Asp	Asp	Arg	Glu	Trp	Gln	Thr	Gln	Leu	Phe	Ala	Leu
				580											585
Leu	Gln	Lys	Gln	Thr	Tyr	Asn	Gln	Val	Glu	Val	Asp	Leu	Phe	Glu	Leu
				595											600
Met	Cys	Lys	Val	Leu	Asp	Gln	Asn	Leu	Phe	Ser	Gln	Val	Asp	Trp	Ala
				610											615
Arg	Asn	Thr	Val	Phe	Phe	Lys	Asp	Leu	Lys	Val	Asp	Asp	Gln	Met	Lys
625					630										635
Leu	Leu	Gln	His	Ser	Trp	Ser	Asp	Met	Leu	Val	Leu	Asp	His	Leu	His
				645											650
His	Arg	Ile	His	Asn	Gly	Leu	Pro	Asp	Glu	Thr	Gln	Leu	Asn	Asn	Gly
				660											665
Gln	Val	Phe	Asn	Leu	Met	Ser	Leu	Gly	Leu	Leu	Gly	Val	Pro	Gln	Leu
				675											680
Gly	Asp	Tyr	Phe	Asn	Glu	Leu	Gln	Asn	Lys	Leu	Gln	Asp	Leu	Lys	Phe
				690											695
Asp	Met	Gly	Asp	Tyr	Val	Cys	Met	Lys	Phe	Leu	Ile	Leu	Leu	Asn	Pro
705					710										715
Ser	Val	Arg	Gly	Ile	Val	Asn	Arg	Lys	Thr	Val	Ser	Glu	Gly	His	Asp
				725											730
Asn	Val	Gln	Ala	Ala	Leu	Leu	Asp	Tyr	Thr	Leu	Thr	Cys	Tyr	Pro	Ser
				740											745
Val	Asn	Asp	Lys	Phe	Arg	Gly	Leu	Val	Asn	Ile	Leu	Pro	Glu	Ile	His
				755											760
Ala	Met	Ala	Val	Arg	Gly	Glu	Asp	His	Leu	Tyr	Thr	Lys	His	Cys	Ala
				770											775
															780

Gly Ser Ala Pro Thr Gln Thr Leu Leu Met Glu Met Leu His Ala Lys  
 785 790 795 800  
 Arg Lys Gly

<210> 4

<211> 3269

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence; note =  
 synthetic construct

<400> 4

ctacgcaaaa	taaaacgtac	atgaaatgtt	attagaaatg	gatcagcaac	aggcgaccgt	60
acagtttata	tcgtcgctga	atatatcgcc	gttcagcatg	cagctggagc	agcagcagca	120
gccctccagt	cccgctctgg	ccgccggtgg	caacagcagc	aacaacgcgg	ccagcggtag	180
caacaacaac	agcgccagcg	gcaacaacac	cagcagcagc	agcaacaaca	acaacaacaa	240
taacaacgac	aatgatgcac	acgttctaac	gaaattcgag	cacgaataca	atgcctacac	300
gttgacagttg	gccggaggcg	gtgggagtg	cagcggcaat	cagcagcacc	acagcaacca	360
cagcaaccac	ggcaaccacc	accagcagca	gcagcaacaa	cagcaacagc	agcagcaaca	420
tcagcagcag	cagcaagaac	actaccagca	gcaacagcaa	cagaatatcg	ccaacaatgc	480
caatcaattc	aactcctcgt	cctactcgta	tatatataat	ttcgattcac	agtatatatt	540
cccgacaggg	taccaggaca	ccacctcctc	acactcgcaa	cagagcggag	gaggcggtag	600
cggcgggcgg	ggcaacctgc	taaacggcag	ctccggcggc	agctccgccc	gcggtaggta	660
catgctgctc	ccccaggcgg	ccagctccag	tggcaataat	ggcaatccga	atgccggcca	720
catgtcctcc	ggttccgtgg	gcaatggcag	cggaggcgct	ggcaatggcg	gagcggggcg	780
caactccggg	cccggcaatc	ccatggggcg	tacgagcgcc	acgcccggac	acggcggcga	840
ggtgatcgac	ttcaagcacc	tgttcgagga	gctttgcccc	gtgtgtggcg	acaaggtag	900
gggctaccac	tacggcctgc	tcacctgcca	gtcctgcaag	ggattcttca	agcgcaccgt	960
gcagaacaag	aaggtctaca	cctgcgtggc	ggagcggtag	tgccacatcg	acaagacgca	1020
gcgcaagcgg	tgtccctact	gccgattcca	gaagtgcctc	gaggtgggca	tgaagctaga	1080
ggctgttcga	gcggatagaa	tgcgtggtgg	acgcaacaaa	ttcggaccca	tgtacaaacg	1140
ggatcgcgcg	cgggaagtgc	aagtgatgcg	gcagcggcag	ttggcgctgc	aagcgctgcg	1200
caactcgatg	ggtccggaca	tcaagccaac	gccgatctcg	ccgggctacc	agcaagcata	1260
tccaaatatg	aacattaagc	aggaaattca	aatacctcag	gtatcctcac	tcacccaatc	1320
tccggactcg	tcgcccagcc	ccatagcaat	tgcgttggga	caggtgaacg	cgagcacggg	1380
cgggtgttata	gccacgcca	tgaacgccgg	cactggcgcc	agtggggggc	gtggtctgaa	1440
cggaccaagt	tccgtgggca	acggcaatag	cagcaacggc	agcagcaacg	gcaacaacaa	1500
cagcagcacg	ggcaacggaa	cgtccggagg	aggaggtagg	aataatgcgg	gcggcggagg	1560
aggaggaacc	aattccaacg	atggcctgca	tcgcaacggc	ggcaatggca	acagcagttg	1620
ccacgaggct	ggaataggat	ctctgcagaa	cacggccgac	tcgaaattgt	gcttcgattc	1680
tggcacacat	ccatcgagca	cagccgacgc	gctaactcag	ccattaagag	tctcaccgat	1740
gattcgtgaa	tttgtgcaat	ctattgacga	tcgggaattg	cagacgcaac	tgtttgccct	1800
gctgcagaag	caaacctaca	accagggtga	agtggatctc	ttcgagctga	tgtgcaaagt	1860
gctcgaccag	aatttggttct	cgcaagtaga	ctgggcacgg	aacaccgtct	tcttcaagga	1920
tctgaaggct	gacgaccaa	tgaagctgct	gcagcattcc	tggtcggaca	tgcttggtct	1980
ggatcacctg	catcatcgaa	tccataacgg	cctgcccgcg	gagacgcaac	tgaacaatgg	2040
tcagggtgttc	aatctgatga	gtctgggttt	gttgggagtg	ccacagctgg	gcgattactt	2100
caacgagctg	cagaacaagc	tgcaggacct	gaaattcgat	atgggcgact	atgtctgcat	2160
gaaattccta	atcctgttga	atccaagtgt	acgggggtatt	gtcaaccgga	agaccgtctc	2220
cgagggacat	gataatgtgc	aagccgcttt	gctggactac	accctcacct	gctatccgtc	2280
agtgaatgac	aaattcagag	ggctagttaa	catcttaccg	gaaatccatg	ccatggccgt	2340
tcgcgggcgg	gatcacctgt	acaccaagca	ctgtgccggc	agtgcgccca	cccaaaccgt	2400
gctcatggag	atgctgcacg	ccaagcgcaa	gggatagagg	ccgggagaac	gtgacacgga	2460
atacttaatc	atttatgaaa	tgtaaataac	aaggcgggaa	ggccctcggg	gcaaccgggt	2520
catggaaggc	gaacgaagga	tacagcagaa	ttccgtatta	tgaatatggg	aatgcatcat	2580
cactactacc	accaactatc	acacctatac	acacacatgc	acacatttgt	tgattcaatg	2640
ttaattatta	ttacgtttac	ggttaggtct	agtttacgtt	taactaatla	attaatttgt	2700
cttaaattaa	ttcgtgtttt	atttgtagtc	cctgataaag	caatttttaa	acacttgaac	2760

```

ctaaacgaga atatgtagta gatgtatgga tttaaattta aatacggcaa ggagaaacac 2820
acttttttag gcattacaaa acaaaagaag catgagaaat ttatttttta tatacctata 2880
tgaatacgat acttatggat acaaatctat atatatTTTT atgtaaattg gcgtactttt 2940
agcgtcctac atatttttta attagaattt gggtatacta tagttttgaa attagtatcg 3000
ttcccacttg aagatcgatt cttgtatttt ttgcgcccaa gtgtcttgca tagtatttgc 3060
gtctaatacta atggcaacaa aaaaaatatt ggaaaatcca tacaagaaa atgaaaacaa 3120
agcaaattta ggtgttcatt gtatgaatgt atgtgtatat tataattgta atttcatcta 3180
agtgtagaaa aacaatgcaa acaactacct acaacaagat aatgaagagc aagaaattat 3240
ataaattaat aaaggtcgtg ttaaaaact 3269

```

&lt;210&gt; 5

&lt;211&gt; 487

&lt;212&gt; PRT

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence; note =  
synthetic construct

&lt;400&gt; 5

```

Met Tyr Thr Gln Arg Met Phe Asp Met Trp Ser Ser Val Thr Ser Lys
 1          5          10          15
Leu Glu Ala His Ala Asn Asn Leu Gly Gln Ser Asn Val Gln Ser Pro
          20          25          30
Ala Gly Gln Asn Asn Ser Ser Gly Ser Ile Lys Ala Gln Ile Glu Ile
          35          40          45
Ile Pro Cys Lys Val Cys Gly Asp Lys Ser Ser Gly Val His Tyr Gly
          50          55          60
Val Ile Thr Cys Glu Gly Cys Lys Gly Phe Phe Arg Arg Ser Gln Ser
          65          70          75          80
Ser Val Val Asn Tyr Gln Cys Pro Arg Asn Lys Gln Cys Val Val Asp
          85          90          95
Arg Val Asn Arg Asn Arg Cys Gln Tyr Cys Arg Leu Gln Lys Cys Leu
          100          105          110
Lys Leu Gly Met Ser Arg Asp Ala Val Lys Phe Gly Arg Met Ser Lys
          115          120          125
Lys Gln Arg Glu Lys Val Glu Asp Glu Val Arg Phe His Arg Ala Gln
          130          135          140
Met Arg Ala Gln Ser Asp Ala Ala Pro Asp Ser Ser Val Tyr Asp Thr
          145          150          155          160
Gln Thr Pro Ser Ser Ser Asp Gln Leu His His Asn Asn Tyr Asn Ser
          165          170          175
Tyr Ser Gly Gly Tyr Ser Asn Asn Glu Val Gly Tyr Gly Ser Pro Tyr
          180          185          190
Gly Tyr Ser Ala Ser Val Thr Pro Gln Gln Thr Met Gln Tyr Asp Ile
          195          200          205
Ser Ala Asp Tyr Val Asp Ser Thr Thr Tyr Glu Pro Arg Ser Thr Ile
          210          215          220
Ile Asp Pro Glu Phe Ile Ser His Ala Asp Gly Asp Ile Asn Asp Val
          225          230          235          240
Leu Ile Lys Thr Leu Ala Glu Ala His Ala Asn Thr Asn Thr Lys Leu
          245          250          255
Glu Ala Val His Asp Met Phe Arg Lys Gln Pro Asp Val Ser Arg Ile
          260          265          270
Leu Tyr Tyr Lys Asn Leu Gly Gln Glu Glu Leu Trp Leu Asp Cys Ala
          275          280          285
Glu Lys Leu Thr Gln Met Ile Gln Asn Ile Ile Glu Phe Ala Lys Leu
          290          295          300
Ile Pro Gly Phe Met Arg Leu Ser Gln Asp Asp Gln Ile Leu Leu Leu
          305          310          315          320

```

Lys Thr Gly Ser Phe Glu Leu Ala Ile Val Arg Met Ser Arg Leu Leu  
 325 330 335  
 Asp Leu Ser Gln Asn Ala Val Leu Tyr Gly Asp Val Met Leu Pro Gln  
 340 345 350  
 Glu Ala Phe Tyr Thr Ser Asp Ser Glu Glu Met Arg Leu Val Ser Arg  
 355 360 365  
 Ile Phe Gln Thr Ala Lys Ser Ile Ala Glu Leu Lys Leu Thr Glu Thr  
 370 375 380  
 Glu Leu Ala Leu Tyr Gln Ser Leu Val Leu Leu Trp Pro Glu Arg Asn  
 385 390 395 400  
 Gly Val Arg Gly Asn Thr Glu Ile Gln Arg Leu Phe Asn Leu Ser Met  
 405 410 415  
 Asn Ala Ile Arg Gln Glu Leu Glu Thr Asn His Ala Pro Leu Lys Gly  
 420 425 430  
 Asp Val Thr Val Leu Asp Thr Leu Leu Asn Asn Ile Pro Asn Phe Arg  
 435 440 445  
 Asp Ile Ser Ile Leu His Met Glu Ser Leu Ser Lys Phe Lys Leu Gln  
 450 455 460  
 His Pro Asn Val Val Phe Pro Ala Leu Tyr Lys Glu Leu Phe Ser Ile  
 465 470 475 480  
 Asp Ser Gln Gln Asp Leu Thr  
 485

&lt;210&gt; 6

&lt;211&gt; 4262

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence; note =  
 synthetic construct

&lt;400&gt; 6

gaattcattc	aactgcaaag	agcagccaaa	ttgcgcatac	gccgcgtatg	gccgtcgggtg	60
tgagtgcccg	tgttcatcag	cggttgcatac	aactgatacc	aagtgtacat	aactacagct	120
acaattgcaa	ctatttcacc	aatcaacggc	agcggcaaca	acatcagcaa	cagcaccggc	180
aaacgtttga	aacgtcacca	aagcttcgca	tttcccacta	ataattatgt	atacgcaacg	240
tatgtttgac	atgtggagca	gcgtcacttc	gaaactggaa	gcacacgcaa	acaatctcgg	300
tcaaagcaac	gtccaatcgc	cggcgggaca	aaacaactcc	agcggttcca	ttaaagctca	360
aattgagata	attccatgca	aagtctgcgg	cgacaagtca	tccggcgtgc	attacggagt	420
gatcacctgc	gagggctgca	agggattctt	tcgaagatcg	cagagctccg	tgggtcaacta	480
ccagtgtccg	cgcaacaagc	aatgtgtggt	ggaccgtggt	aatcgcaacc	gatgtcaata	540
ttgtagactg	caaaagtgcc	taaaactggg	aatgagccgt	gatgctgtaa	agttcggcag	600
gatgtccaag	aagcagcgcg	agaaggtcga	ggacgaggta	cgcttccatc	gggcccagat	660
gcggggcacia	agcgacgcgg	caccggatag	ctccgtatac	gacacacaga	cgccctcgag	720
cagcgaccag	ctgcatcaca	acaattacaa	cagctacagc	ggcggctact	ccaacaacga	780
ggtgggctac	ggcagtcctt	acggatactc	ggcctccgtg	acgccacagc	agaccatgca	840
gtacgacatc	tcggcggact	acgtggacag	caccacctac	gagccgcgca	gtacaataat	900
cgatcccga	tttattagtc	acgcggatgg	cgatatcaac	gatgtgctga	tcaagacgct	960
ggcggaggcg	catgccaaca	caaataccaa	actggaagct	gtgcacgaca	tgttccgaaa	1020
gcagccggat	gtgtcgcgca	ttctctacta	caagaatctg	ggccaagagg	aactctggct	1080
ggactgcgcc	gagaagctta	cacaaatgat	acagaacata	atcgaatttg	ctaagctcat	1140
accgggattc	atgcgcctaa	gtcaggacga	tcagatatta	ctgctgaaga	cgggctcctt	1200
tgagctggcg	attgttcgca	tgtccagact	gcttgatctc	tcacagaacg	cggttctcta	1260
cggcgacgtg	atgctgcccc	aggaggcgtt	ctacacatcc	gactcgggaag	agatgcgtct	1320
ggtgtcgcgc	atcttccaaa	cggccaagtc	gatagccgaa	ctcaaactga	ctgaaaccga	1380
actggcgctg	tatcagagct	tagtgctgct	ctggccagaa	cgcaatggag	tgcgtggtaa	1440
tacggaaata	cagaggcttt	tcaatctgag	catgaatgcg	atccggcagg	agctggaaac	1500
gaatcatgcg	ccgctcaagg	gcgatgtcac	cgtgctggac	acactgctga	acaatatacc	1560
caatttccgc	gatatctcca	tcttgacacat	ggaatcgctg	agcaagttca	agctgcagca	1620
cccgaatgtc	gtttttccgg	cgctgtacaa	ggagctgttc	tcgatagatt	cgcagcagga	1680

```

cctgacataa caagagcagc agccgttcct ggagacgacc gcggacgatg ttgccgagga 1740
tgccggtgcc gccggatgtg tcctgccgcc ggtggcgccc cctgccgggc agcaaccagc 1800
gctgctcgag gactgagggc cgcaggatgt ggcaacaata attatttgag taaacactgc 1860
actgcgcatg cagcagatac aagaacttta tcatgattta agctagcata caaccaagga 1920
tgtgatcctc gcccaaggact cacttaaaaa gaactctatc tatatacata tatatattat 1980
atatgacaga gccggatgacg caaagggaag ggaaaatatt tcaaaaatat tgttaactca 2040
gttaagactt ttgcttcgta gagaaccgaa accgaaaccg attgcatthc gagcaagggg 2100
catcaaactg attttcgagg ttatactata catatataca cacaaacaca cacacacaca 2160
tatatatata tgtaacttcc aaactttcat atcctggccc gagcagatca gatcgtctaa 2220
gtacttaaaa ccaagcgaaa ttctctacac cgcacaaccc aggaccgta gacccaata 2280
attcagttcg gttagtgtta accccagaaa gcccgattcc gatcccgccct aggttgtctt 2340
tgctttacgt tgtaactaaa gtatgtgtat tatatataca gcaaattgtat gtataactat 2400
gtcgtatcgg ttatatgcct aacaacatta ttttttgtaa acaacaaaat cgaatatctc 2460
ggaaaatgtg ttcttataat tatattgatt aatgcaatta caatatattt acaatttacc 2520
gttacgtttt tacattatac ataagacgca agagaaggaa acggaagttt aaggattaga 2580
aagctgaata agaaaaggct taaggacgag ctgagtagca gttaaagtga gcgagaaatc 2640
gaatgaatac cagaaaattt caagcaagca cataaaagta tgcaatattht tgttttaaaaa 2700
caacttttta ttagtttctt aaatataaca taattacgta catacacaca cgtatatata 2760
gggctatata tatctatata tatatatata tacatgatag acaaattcca atccggttcc 2820
aaggtttagt aaaaataaag agaaataaaa cgaaaaacaa aaacttttga tatgaaatcc 2880
tacgcataat taacaacttt tattgtttct aagacttaaa cttaattaaa atggaaacca 2940
aaacagactg acggaccgac cccgacagca tgccacgccc tcccccgccc caccctccac 3000
agatcctggc agaaatttca aaggagtttg atacacaaat cgagaaaaga aattttcaaa 3060
aaaataatat aaagacaagc aaacggcgac ttttttggtt gatacatttg aaaagaatat 3120
acaattaaat atctgactga ctatacaaag acgttacaca cagcataca catacacaca 3180
catacacgca tacacacaca gcttacgata cataaattag ttaaacttag agtaaacaaa 3240
caacaacaaa cacattggat agtaggtgat aattgggtgtg tcttaataaa accttaacct 3300
ctccccgacc cccgcccact tgcttaatac ccaacgcccc aaaaagcccc acatttctac 3360
taaataaaaa gcttaatcaa aacttttttg aaattattca agtgaaaatt tcagcaggca 3420
ggcataaata ttaattaaca ttaattatag caaggaaact tataaataaa atgtatacaa 3480
caaaactaca aaatttaaatt aaattacatt ttgcaaattc cacaqaaaat aaacatgat 3540
tttgcaaatt cacttaaaat cctttccctg aatccaagca aaaatattha cactagctta 3600
catagaactg ggacgaggac atgaatattht caattgagaa aaaaatctat gttaatgtaa 3660
tcgatcgatt tggacatatt taagttcgac atttttggcc ttacaaaaca aaaaacaaaa 3720
agaagaaacc taaagtactt tatatatata caaacatat atacaatata gagaatacaa 3780
aactagtttt aattttatata aagcaaggga gcagctttca aactcaaaac aaaaatatcc 3840
ccgaaaaaaa caacaacttt gttaaaaact gcgcataata aagaaaataa taaacaaagt 3900
taatctataa tataaattga agttaagttg atttgagcgg tcgacaacaa gaacataaat 3960
gtatctttta atgatatatg tattgttaaa tttgtatgct aagtttttag aaaggttaca 4020
tttttaaaaga ataataacaa aagatcgcca actcgacaag gtgtaaaatg agtacattta 4080
aattaaaatt tagcatatat aatgcataaa tattatgtta cgatattht ac atttatataa 4140
aacaacacaa aaacactaaa gaaaaccgaa aaaacagaag tcccatatta aaaatgaaat 4200
aaaatgagca gaacctataa actgataagg gaattctgaa tattaaaaaa aaaaagaaaa 4260
ca

```

&lt;210&gt; 7

&lt;211&gt; 723

&lt;212&gt; PRT

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence; note =  
synthetic construct

&lt;400&gt; 7

```

Met Ser Pro Pro Lys Asn Cys Ala Val Cys Gly Asp Lys Ala Leu Gly
 1          5          10          15
Tyr Asn Phe Asn Ala Val Thr Cys Glu Ser Cys Lys Ala Phe Phe Arg
          20          25          30
Arg Asn Ala Leu Ala Lys Lys Gln Phe Thr Cys Pro Phe Asn Gln Asn
          35          40          45

```

Cys	Asp	Ile	Thr	Val	Val	Thr	Arg	Arg	Phe	Cys	Gln	Lys	Cys	Arg	Leu
50						55				60					
Arg	Lys	Cys	Leu	Asp	Ile	Gly	Met	Lys	Ser	Glu	Asn	Ile	Met	Ser	Glu
65					70					75					80
Glu	Asp	Lys	Leu	Ile	Lys	Arg	Arg	Lys	Ile	Glu	Thr	Asn	Arg	Ala	Lys
				85					90					95	
Arg	Arg	Leu	Met	Glu	Asn	Gly	Thr	Asp	Ala	Cys	Asp	Ala	Asp	Gly	Gly
			100					105					110		
Glu	Glu	Arg	Asp	His	Lys	Ala	Pro	Ala	Asp	Ser	Ser	Ser	Ser	Asn	Leu
			115				120					125			
Asp	His	Tyr	Ser	Gly	Ser	Gln	Asp	Ser	Gln	Ser	Cys	Gly	Ser	Ala	Asp
130						135					140				
Ser	Gly	Ala	Asn	Gly	Cys	Ser	Gly	Arg	Gln	Ala	Ser	Ser	Pro	Gly	Thr
145					150					155					160
Gln	Val	Asn	Pro	Leu	Gln	Met	Thr	Ala	Glu	Lys	Ile	Val	Asp	Gln	Ile
				165					170					175	
Val	Ser	Asp	Pro	Asp	Arg	Ala	Ser	Gln	Ala	Ile	Asn	Arg	Leu	Met	Arg
			180					185					190		
Thr	Gln	Lys	Glu	Ala	Ile	Ser	Val	Met	Glu	Lys	Val	Ile	Ser	Ser	Gln
			195				200						205		
Lys	Asp	Ala	Leu	Arg	Leu	Val	Ser	His	Leu	Ile	Asp	Tyr	Pro	Gly	Asp
210						215					220				
Ala	Leu	Lys	Ile	Ile	Ser	Lys	Phe	Met	Asn	Ser	Pro	Phe	Asn	Ala	Leu
225					230					235					240
Thr	Val	Phe	Thr	Lys	Phe	Met	Ser	Ser	Pro	Thr	Asp	Gly	Val	Glu	Ile
				245					250					255	
Ile	Ser	Lys	Ile	Val	Asp	Ser	Pro	Ala	Asp	Val	Val	Glu	Phe	Met	Gln
			260					265					270		
Asn	Leu	Met	His	Ser	Pro	Glu	Asp	Ala	Ile	Asp	Ile	Met	Asn	Lys	Phe
		275					280					285			
Met	Asn	Thr	Pro	Ala	Glu	Ala	Leu	Arg	Ile	Leu	Asn	Arg	Ile	Leu	Ser
290						295					300				
Gly	Gly	Gly	Ala	Asn	Ala	Ala	Gln	Gln	Thr	Ala	Asp	Arg	Lys	Pro	Leu
305					310					315					320
Leu	Asp	Lys	Glu	Pro	Ala	Val	Lys	Pro	Ala	Ala	Pro	Ala	Glu	Arg	Ala
				325					330					335	
Asp	Thr	Val	Ile	Gln	Ser	Met	Leu	Gly	Asn	Ser	Pro	Pro	Ile	Ser	Pro
			340					345					350		
His	Asp	Ala	Ala	Val	Asp	Leu	Gln	Tyr	His	Ser	Pro	Gly	Val	Gly	Glu
			355				360					365			
Gln	Pro	Ser	Thr	Ser	Ser	Ser	His	Pro	Leu	Pro	Tyr	Ile	Ala	Asn	Ser
370						375					380				
Pro	Asp	Phe	Asp	Leu	Lys	Thr	Phe	Met	Gln	Thr	Asn	Tyr	Asn	Asp	Glu
385					390					395					400
Pro	Ser	Leu	Asp	Ser	Asp	Phe	Ser	Ile	Asn	Ser	Ile	Glu	Ser	Val	Leu
			405						410					415	
Ser	Glu	Val	Ile	Arg	Ile	Glu	Tyr	Gln	Ala	Phe	Asn	Ser	Ile	Gln	Gln
			420					425					430		
Ala	Ala	Ser	Arg	Val	Lys	Glu	Glu	Met	Ser	Tyr	Gly	Thr	Gln	Ser	Thr
			435				440					445			
Tyr	Gly	Gly	Cys	Asn	Ser	Ala	Ala	Asn	Asn	Ser	Gln	Pro	His	Leu	Gln
450						455					460				
Gln	Pro	Ile	Cys	Ala	Pro	Ser	Thr	Gln	Gln	Leu	Asp	Arg	Glu	Leu	Asn
465					470					475					480
Glu	Ala	Glu	Gln	Met	Lys	Leu	Arg	Glu	Leu	Arg	Leu	Ala	Ser	Glu	Ala
			485					490						495	
Leu	Tyr	Asp	Pro	Val	Asp	Glu	Asp	Leu	Ser	Ala	Leu	Met	Met	Gly	Asp
			500					505					510		
Asp	Arg	Ile	Lys	Pro	Asp	Asp	Thr	Arg	His	Asn	Pro	Lys	Leu	Leu	Gln
		515					520					525			

Leu Ile Asn Leu Thr Ala Val Ala Ile Lys Arg Leu Ile Lys Met Ala  
 530 535 540  
 Lys Lys Ile Thr Ala Phe Arg Asp Met Cys Gln Glu Asp Gln Val Ala  
 545 550 555 560  
 Leu Leu Lys Gly Gly Cys Thr Glu Met Met Ile Met Arg Ser Val Met  
 565 570 575  
 Ile Tyr Asp Asp Asp Arg Ala Ala Trp Lys Val Pro His Thr Lys Glu  
 580 585 590  
 Asn Met Gly Asn Ile Arg Thr Asp Leu Leu Lys Phe Ala Glu Gly Asn  
 595 600 605  
 Ile Tyr Glu Glu His Gln Lys Phe Ile Thr Thr Phe Asp Glu Lys Trp  
 610 615 620  
 Arg Met Asp Glu Asn Ile Ile Leu Ile Met Cys Ala Ile Val Leu Phe  
 625 630 635 640  
 Thr Ser Ala Arg Ser Arg Val Ile His Lys Asp Val Ile Arg Leu Glu  
 645 650 655  
 Gln Asn Ser Tyr Tyr Tyr Leu Leu Arg Arg Tyr Leu Glu Ser Val Tyr  
 660 665 670  
 Ser Gly Cys Glu Ala Arg Asn Ala Phe Ile Lys Leu Ile Gln Lys Ile  
 675 680 685  
 Ser Asp Val Glu Arg Leu Asn Lys Phe Ile Ile Asn Val Tyr Leu Asn  
 690 695 700  
 Val Asn Pro Ser Gln Val Glu Pro Leu Leu Arg Glu Ile Phe Asp Leu  
 705 710 715 720  
 Lys Asn His

&lt;210&gt; 8

&lt;211&gt; 2832

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence; note =  
 synthetic construct

&lt;400&gt; 8

gttattggga	ttggcctgga	gcactcggac	ggacagtaat	tcattaaaat	atgtggtgat	60
aacgcgagct	gccgaatctg	cgtgcaattc	gtgcgtttga	cgtgggtact	aactgctatg	120
ctgtcgcgcg	gacagttggt	ctgatacgca	gagttcctgc	ctcaccacac	acgaccacct	180
ccattaaaac	cagccacccc	ccccagcgcc	tcctccaccg	acagcagctg	ctccaccgca	240
ccaccaggag	aggggcaatt	aaaaaatcaa	tcagagggcc	ctaattgaaa	gctgccaccg	300
tcgaaatgtc	gccgccgaag	aactgcgcgg	tgtgcgggga	caaggctctg	ggctacaact	360
tcaatgcggt	cacctgcgag	agctgcaagg	cgttcttccg	acggaacgcg	ctggccaaga	420
agcagttcac	ctgccccttc	aacccaaaact	gcgacatcac	tgtggtcact	cgacgcttct	480
gccagaaatg	ccgcctgcgc	aagtgcctgg	atatcgggat	gaagagtga	aacattatgt	540
ccgaggagga	caagctgata	aagcggcgca	agatcgagac	caaccgggcc	aagcgacgcc	600
tcatggagaa	cggcacggat	gcgtgcgacg	ccgatggcgg	cgaggaaagg	gatcacaagg	660
cgccggcgga	tagcagcagc	agcaaccttg	accactactc	ggggtcacag	gactcgcaga	720
gctgcggctc	ggcggacagc	ggggccaatg	ggtgctccgg	cagacaggcc	agttcgccgg	780
gcacacaggt	caatccgctt	cagatgacgg	ccgagaagat	agtcgaccag	atcgtatccg	840
acccggatcg	agcctcgcag	gccatcaacc	ggttgatgcg	cacgcagaaa	gaggctatat	900
cggtgatgga	gaaggtaatc	agctcacaaa	aggacgcctt	aaggctggtg	tcgcatttga	960
tcgactatcc	aggcgacgca	ctcaagatca	tttcaaagtt	tatgaactcg	ccctttaacg	1020
cgctgacagt	attcaccaaa	ttcatgagct	caccacagga	cggcgttgaa	attatctcaa	1080
agatagttga	ttcgcccgcg	gacgtggtgg	agttcatgca	gaacttgatg	cactcgccag	1140
aggacgccat	cgatataatg	aacaagttca	tgaatacccc	agcggaggcg	ctgcgcattc	1200
ttaaccgaat	cctaagcggc	ggaggagcga	acgcagccca	gcagacagca	gaccgcaagc	1260
cattgctgga	caaggagccg	gcggtgaagc	ctgcagcgcc	agcggagcga	gctgatactg	1320
tcattcaaag	catgctgggc	aacagtccgc	caatttcgcc	acatgatgct	gccgtggatc	1380
tgcagtacca	ctcgcccggg	gtcgggggagc	agcccagtac	atcgagtagc	cacccttgc	1440

```

cttacatagc caactcgccg gacttcgata tgaagacctt catgcagacc aactacaacg 1500
acgagcccag tctggacagt gatttttagca ttaactcaat cgaatcggtg ctatccgagg 1560
tgatccgcat tgagtaccag gccttcaata gcatacaaca agcggcatcg cgcgtaaagg 1620
aggagatgtc ctacggcact cagtctacgt acggtggatg caattcggct gcaaacaata 1680
gccagccgca cctgcagcaa cccatctgcg ccccatccac ccagcagttg gatcgcgagc 1740
taaacgaggg ggagcaaagt aagctgcggg agctgcgact ggccagcgag gctctttatg 1800
atcccgtgga cgaggacctc agcgccctga tgatgggcca tgatcgcat aagcccgcag 1860
acactcgcca caacccaaag ctattgcagc tgatcaatct gacggcggtg gccatcaagc 1920
ggcttatcaa aatggccaag aagattacag cattccgtga catgtgccag gaggaccagg 1980
tggccctact caaaggtggc tgcacagaaa tgatgataat gcgctccgta atgatttacg 2040
acgacgatcg cgccgcctgg aaggtacccc ataccaaaga gaacatgggc aacatacgca 2100
ctgacctgct caagtttgcc gaaggcaata tctacgagga gcaccaaag ttcatacaca 2160
cgtttgacga gaagtggcgc atggacgaga acataatcct gatcatgtgt gccattgtcc 2220
tttttacctc ggctcgatcg cgagtgatac acaaagacgt gattagattg gaacagaatt 2280
cctactatta tcttctgcca agatatctgg agagtgttta ttctggctgt gaggcgagaa 2340
acgcgtttat caagctaata caaaagattt cagatgtgga gcgtctgaac aagttcataa 2400
ttaatgtcta tttgaatgtt aaccatccc aggtggagcc cttgctgcgt gaaatattcg 2460
at ttgaaaaa t cactagaca accgatgcgt gtcgggcatt taatgcctat gttgatgccc 2520
aatgatgaat ggtcaacaag ctgtagttgt tgttgttgtt gatgtctgtt ttatcttgte 2580
gcttgtaatg ttagatttta atcgaatgtg attgttagat ttgcatatac tgcatagatt 2640
ttatatcttct acatcaaaga gagcatattt aggataccaa gtgcaaagca acacaatcta 2700
tatgtaatgt acaccgttta cctagtttca aataaactag acgataatgc aataactaac 2760
ttggaagcgt gggttctgtg caaaaaggaa aaaagacaaa aaaaataaac tgactttgag 2820
aaccagtggg aa 2832

```

&lt;210&gt; 9

&lt;211&gt; 704

&lt;212&gt; PRT

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence; note =  
synthetic construct

&lt;400&gt; 9

```

Met Met Lys His Pro Gln Asp Leu Ser Val Thr Asp Asp Gln Gln Leu
 1             5             10             15
Met Lys Val Asn Lys Val Glu Lys Met Glu Gln Glu Leu His Asp Pro
             20             25             30
Glu Ser Glu Ser His Ile Met His Ala Asp Ala Leu Ala Ser Ala Tyr
             35             40             45
Pro Ala Ala Ser Gln Pro His Ser Pro Ile Gly Leu Ala Leu Ser Pro
             50             55             60
Asn Gly Gly Gly Leu Gly Leu Ser Asn Ser Ser Asn Gln Ser Ser Glu
65             70             75             80
Asn Phe Ala Leu Cys Asn Gly Asn Gly Asn Ala Gly Ser Ala Gly Gly
             85             90             95
Gly Ser Ala Ser Ser Gly Ser Asn Asn Asn Ser Met Phe Ser Pro
             100            105            110
Asn Asn Asn Leu Ser Gly Ser Gly Ser Gly Thr Asn Ser Ser Gln Gln
             115            120            125
Gln Leu Gln Gln Gln Gln Gln Gln Ser Pro Thr Val Cys Ala Ile
             130            135            140
Cys Gly Asp Arg Ala Thr Gly Lys His Tyr Gly Ala Ser Ser Cys Asp
145            150            155            160
Gly Cys Lys Gly Phe Phe Arg Arg Ser Val Arg Lys Asn His Gln Tyr
             165            170            175
Thr Cys Arg Phe Ala Arg Asn Cys Val Val Asp Lys Asp Lys Arg Asn
             180            185            190
Gln Cys Arg Tyr Cys Arg Leu Arg Lys Cys Phe Lys Ala Gly Met Lys
             195            200            205

```

Lys	Glu	Ala	Val	Gln	Asn	Glu	Arg	Asp	Arg	Ile	Ser	Cys	Arg	Arg	Thr
210						215					220				
Ser	Asn	Asp	Asp	Pro	Asp	Pro	Gly	Asn	Gly	Leu	Ser	Val	Ile	Ser	Leu
225					230					235					240
Val	Lys	Ala	Glu	Asn	Glu	Ser	Arg	Gln	Ser	Lys	Ala	Gly	Ala	Ala	Met
				245					250					255	
Glu	Pro	Asn	Ile	Asn	Glu	Asp	Leu	Ser	Asn	Lys	Gln	Phe	Ala	Ser	Ile
			260					265					270		
Asn	Asp	Val	Cys	Glu	Ser	Met	Lys	Gln	Gln	Leu	Leu	Thr	Leu	Val	Glu
		275					280					285			
Trp	Ala	Lys	Gln	Ile	Pro	Ala	Phe	Asn	Glu	Leu	Gln	Leu	Asp	Asp	Gln
290						295					300				
Val	Ala	Leu	Leu	Arg	Ala	His	Ala	Gly	Glu	His	Leu	Leu	Leu	Gly	Leu
305					310					315					320
Ser	Arg	Arg	Ser	Met	His	Leu	Lys	Asp	Val	Leu	Leu	Leu	Ser	Asn	Asn
				325					330					335	
Cys	Val	Ile	Thr	Arg	His	Cys	Pro	Asp	Pro	Leu	Val	Ser	Pro	Asn	Leu
			340					345					350		
Asp	Ile	Ser	Arg	Ile	Gly	Ala	Arg	Ile	Ile	Asp	Glu	Leu	Val	Thr	Val
		355					360					365			
Met	Lys	Asp	Val	Gly	Ile	Asp	Asp	Thr	Glu	Phe	Ala	Cys	Ile	Lys	Ala
370						375					380				
Leu	Val	Phe	Phe	Asp	Pro	Asn	Ala	Lys	Gly	Leu	Asn	Glu	Pro	His	Arg
385					390					395					400
Ile	Lys	Ser	Leu	Arg	His	Gln	Ile	Leu	Asn	Asn	Leu	Glu	Asp	Tyr	Ile
				405					410					415	
Ser	Asp	Arg	Gln	Tyr	Glu	Ser	Arg	Gly	Arg	Phe	Gly	Glu	Ile	Leu	Leu
			420					425					430		
Ile	Leu	Pro	Val	Leu	Gln	Ser	Ile	Thr	Trp	Gln	Met	Ile	Glu	Gln	Ile
		435					440					445			
Gln	Phe	Ala	Lys	Ile	Phe	Gly	Val	Ala	His	Ile	Asp	Ser	Leu	Leu	Gln
450						455					460				
Glu	Met	Leu	Leu	Gly	Gly	Glu	Leu	Ala	Asp	Asn	Pro	Leu	Pro	Leu	Ser
465					470					475					480
Pro	Pro	Asn	Gln	Ser	Asn	Asp	Tyr	Gln	Ser	Pro	Thr	His	Thr	Gly	Asn
				485					490					495	
Met	Glu	Gly	Gly	Asn	Gln	Val	Asn	Ser	Ser	Leu	Asp	Ser	Leu	Ala	Thr
			500					505					510		
Ser	Gly	Gly	Pro	Gly	Ser	His	Ser	Leu	Asp	Leu	Glu	Val	Gln	His	Ile
		515					520					525			
Gln	Ala	Leu	Ile	Glu	Ala	Asn	Ser	Ala	Asp	Asp	Ser	Phe	Arg	Ala	Tyr
530						535					540				
Ala	Ala	Ser	Thr	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Val	Ser	Ser	Ser	Ser
545					550					555					560
Ser	Ala	Pro	Ala	Ser	Val	Ala	Pro	Ala	Ser	Ile	Ser	Pro	Pro	Leu	Asn
				565					570					575	
Ser	Pro	Lys	Ser	Gln	His	Gln	His	Gln	Gln	His	Ala	Thr	His	Gln	Gln
			580					585					590		
Gln	Gln	Glu	Ser	Ser	Tyr	Leu	Asp	Met	Pro	Val	Lys	His	Tyr	Asn	Gly
		595					600					605			
Ser	Arg	Ser	Gly	Pro	Leu	Pro	Thr	Gln	His	Ser	Pro	Gln	Arg	Met	His
610						615					620				
Pro	Tyr	Gln	Arg	Ala	Val	Ala	Ser	Pro	Val	Glu	Val	Ser	Ser	Gly	Gly
625					630					635					640
Gly	Gly	Leu	Gly	Leu	Arg	Asn	Pro	Ala	Asp	Ile	Thr	Leu	Asn	Glu	Tyr
				645					650					655	
Asn	Arg	Ser	Glu	Gly	Ser	Ser	Ala	Glu	Glu	Leu	Leu	Arg	Arg	Thr	Pro
			660					665					670		
Leu	Lys	Ile	Arg	Ala	Pro	Glu	Met	Leu	Thr	Ala	Pro	Ala	Gly	Tyr	Gly
		675					680					685			

Thr Glu Pro Cys Arg Met Thr Leu Lys Gln Glu Pro Glu Thr Gly Tyr  
 690 695 700

<210> 10

<211> 3248

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence; note =  
 synthetic construct

<400> 10

agttgaattc	cagtgacggt	ggaagaaaca	actgcaaaag	gcaaaaacaa	agacaatggt	60
tataagctgt	atattccgct	ttgattgata	taaatgaata	tatgcagtgc	gccagttata	120
caactgccct	gcaaaagtca	ctcattaaat	aaaaaacgcc	cgagatgaat	ttcacagcgg	180
cggcaacaag	tgcaataata	gtaaaaaatc	aaaagccaaa	caacgaaatc	tctcccaaaa	240
aaacgaagaa	gcgtgtcgcg	gtgcacaaaa	gaaaacaaaa	atagaaaaat	acacaacaaa	300
ataatacggg	gaaacgttaa	ttataacgag	ccacaaaatc	gcataaagaa	atcaacaagt	360
gtgtgtctgc	ctttttttcc	atattcgctt	tcattcatgc	ggtcaactca	acaataacaa	420
ctcaaaatag	caacaacaac	aataacaata	tcaacaagag	cagcagcagt	cgctgataaa	480
agccctgcag	ctaaaacaac	aacaaaacaa	caaagatagt	tagaaagaac	atcgtctggc	540
cattgagctt	taattgccgg	tcattacttc	attactatgt	gattggatct	tcccgaccca	600
cttgtaaata	aaaagtaaaa	atactgggta	tgaagcatga	tgaagcatcc	gcaggatctg	660
agtgtcacgg	atgaccagca	gttaatgaag	gtgaacaagg	tggagaagat	ggagcaggag	720
ttgcacgacc	ccgaatcgga	gagccacata	atgcacgcgg	atgccctggc	ctctgcctat	780
ccggctgcct	cgcagcccca	cagtcgcgatc	ggcctcgccc	tcagcccca	tggcgggtggg	840
ctgggactga	gcaacagtag	caaccagagc	agcgagaact	ttgcgctctg	caacggaaac	900
ggaaatgcgg	gcagcgcagg	aggcgggaagt	gccagcagtg	gcagcaacaa	caacaacagc	960
atgttctcac	ccaacaacaa	cttgagcggg	agcgggaagt	ggactaacag	cagtcagcag	1020
caattgcagc	agcaacaaca	acagcaatca	ccgacgggtc	gcgccatttg	tggagatcgg	1080
gcgacgggca	aacattatgg	agcctccagc	tgcgacggct	gcaaaggatt	cttcaggagg	1140
agtgtcagga	aaaatcatca	gtacacttgc	agatttgcgc	gaaactgcgt	tgtggacaag	1200
gacaaacgga	atcagtgccg	ctactgccgg	ctgaggaagt	gcttcaaggc	gggcatgaag	1260
aaggaggcgg	tgcaaaacga	gcgggatcgc	attagctgcc	gccgcacctc	caatgacgac	1320
ccggatccgg	gcaatgggct	gtctgtgatt	tccttggtta	aggcggagaa	tgagtgcgct	1380
cagtcgaagg	caggcgctgc	catggagcca	aacattaacg	aggacctctc	caacaagcag	1440
ttcgcgagca	tcaacgatgt	ctgcgagtcg	atgaagcagc	agctgctgac	cctgggtggaa	1500
tgggctaagc	agattccggc	ctttaacgag	ctgcagctgg	atgaccagg	ggcactgcta	1560
cgcgcccatg	ctggcgagca	tttgctcctc	ggcctgtctc	gtcgttcgat	gcacttgaag	1620
gatgttctcc	tgctgagcaa	caattgtgtg	atcacaaggc	actgtccaga	tccccttggtg	1680
tcgccgaatt	tggacatctc	ccggatccgg	gcccgtatca	tcgatgaact	ggtgacggtc	1740
atgaaggatg	tgggtatcga	tgacactgaa	ttcgcttgca	tcaaggccct	agtcttcttc	1800
gatcccaatg	ccaagggtct	taatgaaccg	catcgcatca	aatcgctacg	gcacagata	1860
ctcaataatc	tcgaggacta	catatcagat	cggcaataacg	agtcgcgcgg	tcgctttggc	1920
gagattctgc	tcattcctgcc	ggttctgcag	tctattacct	ggcagatgat	cgagcagatc	1980
cagtttgcca	agatctttgg	agtggcccac	attgattcat	tactgcagga	aatgttggtg	2040
ggaggagagt	tggccgacaa	tcctctgccg	ctatcgccgc	ccaatcagtc	aaatgactac	2100
cagagtccca	cccacacagg	caacatggag	ggcggtaatc	aagttaactc	ctctctggac	2160
tcgctggcca	cgtccggtgg	tcctggctcg	catagtctgg	acctggagg	gcagcacatt	2220
caggctctta	tcgaggcgaa	cagtgcggat	gattccttcc	gggcctacgc	ggccagcact	2280
gcagcggcag	ccgctgcagc	cgtctcgtec	tcctcctctg	cacccgcac	cgttgctcca	2340
gcctcgatct	ctcctccgct	caacagcccc	aagtcacaac	atcaacatca	gcaacatgcg	2400
acgcacgac	aacaacagga	gagctcctac	ttggacatgc	ccgtcaagca	ctacaatggc	2460
agtcgggtccg	gaccgctgcc	aacacagcac	agtcaccaga	ggatgcaccc	ctaccaaaga	2520
gcagtcgcct	cgccggctga	agtgtccagc	gggggcggcg	gattgggtct	gcgcaatcct	2580
gccgatatta	cgctcaacga	gtacaaccgg	agcgagggtta	gcagtgccga	ggagctgctg	2640
cgacgaactc	cactgaagat	ccgggctccc	gagatgctaa	ccgcacccgc	tggttatgga	2700
acggaaccct	gtcgcatgac	acttaaacag	gagccagaga	ctggttacta	gaagaataac	2760
gaacgggtgca	atatgcagtt	tgcaatagga	caccctttaa	gcacacaacc	catacacata	2820
caggccctct	cttgctgtac	tccccaccaa	gtgctatata	gagatgaaat	tgaaatgaag	2880

```

aacttactta attggttatgc cttgaaccat tttgatactt tttattagtc ctaagtaggt 2940
atatttggaat ttgttgctta atttttaaatg tttaacgcag ttgcaatata ttttttgagt 3000
catatttttgc tcaagaagtt tattatatac aattatacta tatatataca ccatttagca 3060
tgtactgagt ttgttggtta tttgggttatc ttatacttgt gcgtggatca caaaacattc 3120
atataaggcc atgcaatata ttgttttagg ttaggggtgt gtctagatta tgctgaaagt 3180
gtaatatata tttaatttta aacaaagaac tattttttata tgaatatgta taatatacaa 3240
actatttc 3248

```

&lt;210&gt; 11

&lt;211&gt; 556

&lt;212&gt; PRT

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence; note =  
synthetic construct

&lt;400&gt; 11

```

Met Asp Glu Asp Cys Phe Pro Pro Leu Ser Gly Gly Trp Ser Ala Ser
1      5      10      15
Pro Pro Ala Pro Ser Gln Leu Gln Gln Leu His Thr Leu Gln Ser Gln
20      25      30
Ala Gln Met Ser His Pro Asn Ser Ser Asn Asn Ser Ser Asn Asn Ala
35      40      45
Gly Asn Ser His Asn Asn Ser Gly Gly Tyr Asn Tyr His Gly His Phe
50      55      60
Asn Ala Ile Asn Ala Ser Ala Asn Leu Ser Pro Ser Ser Ser Ala Ser
65      70      75      80
Ser Leu Tyr Glu Tyr Asn Gly Val Ser Ala Ala Asp Asn Phe Tyr Gly
85      90      95
Gln Gln Gln Gln Gln Gln Gln Gln Ser Tyr Gln Gln His Asn Tyr Asn
100     105     110
Ser His Asn Gly Glu Arg Tyr Ser Leu Pro Thr Phe Pro Thr Ile Ser
115     120     125
Glu Leu Ala Ala Ala Thr Ala Ala Val Glu Ala Ala Ala Ala Ala Thr
130     135     140
Val Ser Ser Pro Ser Val Gly Gly Pro Pro Pro Val Arg Arg Ala Ser
145     150     155     160
Leu Pro Val Gln Arg Thr Val Ser Pro Ala Gly Ser Thr Ala Gln Ser
165     170     175
Pro Lys Leu Ala Lys Ile Thr Leu Asn Gln Arg His Ser His Ala His
180     185     190
Ala His Ala Leu Gln Leu Asn Ser Ala Pro Asn Ser Ala Ala Ser Ser
195     200     205
Pro Ala Ser Ala Asp Leu Gln Ala Gly Arg Leu Leu Gln Ala Pro Ser
210     215     220
Gln Leu Cys Ala Val Cys Gly Asp Thr Ala Ala Cys Gln His Tyr Gly
225     230     235     240
Val Arg Thr Cys Glu Gly Cys Lys Gly Phe Phe Lys Arg Thr Val Gln
245     250     255
Lys Gly Ser Lys Tyr Val Cys Leu Ala Asp Lys Asn Cys Pro Val Asp
260     265     270
Lys Arg Arg Arg Asn Arg Cys Gln Phe Cys Arg Phe Gln Lys Cys Leu
275     280     285
Val Val Gly Met Val Lys Glu Val Val Arg Thr Asp Ser Leu Lys Gly
290     295     300
Arg Arg Gly Arg Leu Pro Ser Lys Pro Lys Ser Pro Gln Glu Ser Pro
305     310     315     320
Pro Ser Pro Pro Ile Ser Leu Ile Thr Ala Leu Val Arg Ser His Val
325     330     335

```

Asp Thr Thr Pro Asp Pro Ser Cys Leu Asp Tyr Ser His Tyr Glu Glu  
 340 345 350  
 Gln Ser Met Ser Glu Ala Asp Lys Val Gln Gln Phe Tyr Gln Leu Leu  
 355 360 365  
 Thr Ser Ser Val Asp Val Ile Lys Gln Phe Ala Glu Lys Ile Pro Gly  
 370 375 380  
 Tyr Phe Asp Leu Leu Pro Glu Asp Gln Glu Leu Leu Phe Gln Ser Ala  
 385 390 395 400  
 Ser Leu Glu Leu Phe Val Leu Arg Leu Ala Tyr Arg Ala Arg Ile Asp  
 405 410 415  
 Asp Thr Lys Leu Ile Phe Cys Asn Gly Thr Val Leu His Arg Thr Gln  
 420 425 430  
 Cys Leu Arg Ser Phe Gly Glu Trp Leu Asn Asp Ile Met Glu Phe Ser  
 435 440 445  
 Arg Ser Leu His Asn Leu Glu Ile Asp Ile Ser Ala Phe Ala Cys Leu  
 450 455 460  
 Cys Ala Leu Thr Leu Ile Thr Glu Arg His Gly Leu Arg Glu Pro Lys  
 465 470 475 480  
 Lys Val Glu Gln Leu Gln Met Lys Ile Ile Gly Ser Leu Arg Asp His  
 485 490 495  
 Val Thr Tyr Asn Ala Glu Ala Gln Lys Lys Gln His Tyr Phe Ser Arg  
 500 505 510  
 Leu Leu Gly Lys Leu Pro Glu Leu Arg Ser Leu Ser Val Gln Gly Leu  
 515 520 525  
 Gln Arg Ile Phe Tyr Leu Lys Leu Glu Asp Leu Val Pro Ala Pro Ala  
 530 535 540  
 Leu Ile Glu Asn Met Phe Val Thr Thr Leu Pro Phe  
 545 550 555

&lt;210&gt; 12

&lt;211&gt; 5181

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence; note =  
 synthetic construct

&lt;400&gt; 12

ctcgccatt	ggagggcccc	tgctctgtgg	cagcagcttg	cccagcttcc	aggagaccta	60
ctccttgaag	tacaacagca	gcagcggtag	cagccccag	caggcgctcct	cctcctccac	120
cgccgcccc	acgcccactg	accaggtgct	gacctcaag	atggacgagg	actgcttccc	180
gcctctgtcc	ggcggctgga	gtgccagtcc	gcccgcccc	tcccagctcc	agcagctgca	240
cacctgcag	tctcaggccc	agatgtcgca	tcccaacagc	agcaacaaca	gcagcaacaa	300
cgcgggcaac	agccacaaca	acagtggggg	ctacaactac	cacggccact	tcaatgccat	360
caatgccagc	gccaatctgt	cgcccagctc	ctcggccagt	tccctctacg	aatataatgg	420
tgtttccgca	gcggacaact	tctacggaca	acagcagcag	cagcaacagc	aaagctatca	480
gcaacataac	tacaactcgc	acaatggcga	gcgttactcg	ctgcccacgt	ttcccacgat	540
ttcggagctg	gctgcggcca	ctgctgctgt	cgaagctgcg	gcggcggcca	cagtctctctc	600
cccttcggtg	ggcgggtccgc	cgccagtacg	ccgagcatcg	ctgcccgttc	agcgaaccgt	660
ttcgccagcc	ggctccacgg	cgcagagccc	caagctggcc	aagatcacac	tgaaccagcg	720
gcactcccat	gcccattgcc	atgccctaca	gctcaactcg	gcacccaatt	cggcggcaag	780
ttcgccagcg	agtgcggatc	tcgaggcggg	ccgtttgctc	caggctccgt	cgcagctgtg	840
tgccgtttgt	ggcgacaccg	ccgcctgcc	gcattatgga	gtgcgaacct	gcgagggatg	900
caagggattc	ttcaagcgga	ccgtgcagaa	gggctccaag	tatgtctgcc	tagcggacaa	960
gaattgcccg	gtggacaaga	ggcgccgcaa	ccgttgccag	ttctgcccgt	tccagaagtg	1020
cctggtcgta	ggcatggtca	aggaagtggg	gcgcacggac	tcgttgaagg	gtcgccgcgg	1080
gagactgccc	tcaaaaccga	aatcgcccc	ggagtcgcca	ccatcaccac	ccatctcggt	1140
gatcacggcc	ctggttcgca	gccatgtcga	cacgactccg	gatccctcgt	gcctggacta	1200
cagccactat	gaggagcagt	cgatgagcga	ggcagataag	gtgcaacagt	tttaccagct	1260
gctgaccagc	tccgtggacg	tgatcaagca	gttcgcccag	aagattcccc	gctacttcga	1320

tctcctgccg	gaggatcagg	agctgctctt	ccagagcgca	tcgctggaac	tgttcgtcct	1380
gcggctggcc	tatcgcgcca	ggatcgatga	caccaagctg	atcttctgca	acggcacggg	1440
gctccaccgc	acccagtgcc	tgcgctcctt	cggcgagtgg	ctcaacgaca	tcattggagt	1500
cagccgcagc	ctgcacaacc	tggagatcga	catctccgcc	ttcgccctgcc	tctgtgccct	1560
aaccctgac	acagaacgcc	atggcctgcy	ggagccgaag	aaggtggagc	agctccagat	1620
gaagatcatt	ggcagtctgc	gcgaccacgt	cacctacaat	gccgaggccc	agaagaagca	1680
gcactacttc	agccgcctgc	tgggcaagct	gccggagctg	aggtccctga	gtgtccaggg	1740
actgcagagg	atcttctacc	tgaagctgga	ggacctgggt	cccgcgccag	ctctcatcga	1800
gaacatgttc	gtcaccacat	tgccttctta	gagcgatca	tcaagcgtat	catcacaact	1860
tgcttcctta	aactagcccc	taagttatgc	ctcctaggat	atacagagaa	aggaccccat	1920
aggacggacg	caactagctt	tagtagaacc	ctgaaataaa	taaatctcac	aacagcaaaa	1980
acaaaaccga	accgaacaga	aatgaagcga	atagcagacc	caggccatat	ctttagtgtg	2040
gagctaggta	gttagccgga	cagccccggc	tccttcgata	attacggaca	tgcataatttg	2100
agaggggggt	tccagtgcac	agcctatggc	tcctgcgtga	ctcgtcagca	ccgcgagctc	2160
caacttggtg	acgttaattg	ttaaattggt	taatttcaac	tgtcaaaacc	ggaatcaacg	2220
gccgggcacg	caatggcaac	actttctatc	cccggacttc	gaagcctgct	caacattcgg	2280
cactacggac	ggacaaacaa	cggacagaaa	cagaactcac	tcttgctctc	ttgccttttg	2340
ctaacttcta	gtcaattgat	ttaggcgaat	caaataaata	aataaataaa	ataagggcgt	2400
gcagcagtag	tgttatataa	tttctatgcc	agaccccgag	ggttctcttc	aaggaaatcc	2460
cccaatgagt	tgcacaaatt	gggataaagt	acgatagcct	attattctta	tatttctttt	2520
aaaagctcga	agatagatga	gaactgtgtg	gaaatccact	atcatatcat	atagttgcta	2580
taagccgtgc	ttgccctaag	ctaagttaga	cccgcataaa	gttgatagcc	caaccaagta	2640
tttcgggttat	ttcctagact	aaggtcctaa	tagttatagg	ctaagactat	tctgttcgat	2700
ttatcaatgc	accaaacagt	gcacaatgag	agtataagta	ccttcttggt	atgattgtgt	2760
ctgacacaga	gagagttgca	cacaagcaca	caaactagcc	gataagttac	taaatacgat	2820
ctaataatcta	atatatataa	tataatataa	tatatataag	tccaagtatt	cggaaatcca	2880
agaacccttg	cataaccgca	gttcgtacgt	tccaaacgag	aaaagaactt	tattttaatcc	2940
tagaccactc	catctaagtt	ctcaaagaat	cgtatgtgga	tcggtggatc	tgtctctcta	3000
tatatgtgtg	tgtgttatct	cgatagaaaa	cccctctatg	tgattttgtg	atagattggc	3060
attgaactct	atatatattat	atatatatgt	ctataatata	tatacacgca	taaatatata	3120
tttttatgtc	taacttttgt	atggtttatt	ttatacgtac	cacttttctt	tgataacaaa	3180
aagtaaaaaa	ctcgtttagat	agcaaataat	tcaaagggtat	gttacgagga	cttttcaaag	3240
taccagtctt	tagcgacttt	ccaattaacg	ttcgtattaa	cgaaagacag	attttctatg	3300
tgttaaattg	aagacttcta	taactataac	taaatgcaag	ctaagagcaa	aaacacaaat	3360
ccacaaatcc	ccaaagtga	taacatatct	cttcaagctt	tcgagtgcac	ggaacacgta	3420
gaaccgaaac	ccaagtgtta	ctaaatccat	ttaataatcg	gcaagccggg	ggcgtcggcg	3480
tgggttaatac	gttctcatta	cctatacaat	ttagatagat	cattattaaa	ttattgtaca	3540
tgtagcacat	gaaatgttcg	acaactagat	tttgtaccat	cttaaagaag	aacctaggcc	3600
aagctaaact	aagtataaac	tatgatctgc	atgcggctga	gctgtagcta	tgagaaatat	3660
acctgcgtgg	atctaagtga	aatgggacac	tttgaattta	gatatgaaac	gttctaaacg	3720
cgacgtacta	actctcccaa	ctgcgaactc	taccaattaa	gagaaattcc	cagaaaatgt	3780
gtcaggattt	caaagcgtcc	catctcactt	gaacccaccc	aatcaacaaa	tacaaatcct	3840
agggaaagtg	agaggttcag	caaccataga	gcaatatttc	ataagaaaac	gcaccttaaa	3900
ttaccgaaaa	acatagatta	acctgatctt	gtaacgtttg	ggagcgataa	taagccagga	3960
ttaaacagga	acagttaggt	gaccaaatca	gttcgaaacg	agatgataga	taggttcggg	4020
ttcgaaaccc	taaacgcgat	gccatttttag	ccgttacaac	attggatatc	aacctatgcac	4080
atgaatatga	atatgaatat	gaatattata	gagatatatc	tagctatagg	aacctacttt	4140
gtacctacac	gacatggaaa	catcaaacct	acatgcatat	ttacacacat	atattttgaa	4200
tagagcgacg	actttttacaa	gttgcggtaca	aagctatagc	tatagcttga	tatggccatc	4260
ccagagcgag	catatacata	tatttttgggt	tattgttctt	ttgtaatttt	ataaatgcat	4320
acataatttat	tgtactacgt	gaatgtcaag	tgtggattca	tattttttgag	atacagctac	4380
aaaacgaaac	aaaagaaaat	aaaacaaaac	agaagagtaa	acgtgaaatt	tttcgatgaa	4440
acaatttttaa	atgagaactt	tttaatatgt	ctattaaagg	atatacatat	acacactaac	4500
atacatatat	attttactat	gtaacggata	gaattaagct	agatgcagcg	cataaagctt	4560
tatacaacaa	attgaaaagc	aacagaagaa	attggcacia	attaaattta	tatagcataa	4620
ttagacgtcc	ttcgcaagat	aatgttatct	gtaataagag	cgtcaatcgg	tacatcgggc	4680
gctattttccc	actacacccc	caaccacaca	atagataacc	taagctatgt	atgtacatta	4740
gctatgtata	tccagcccac	ttatgcgcct	actactagaa	atgcagaaag	cagaaagaga	4800
ggtgaaacct	atagacgcta	tcacaaatgt	ctatctgata	gacatcggta	ctaccaatgc	4860
tatattgcca	gttgtgtaat	ttactcttat	ttgatcggtt	catttaccag	ttaagaaccc	4920
aatcatata	agtgttatga	tggagaagact	ataacttgca	attcaattaa	ctctgcaata	4980

cgataacaag	caaagcgaat	catttcattt	cgattttaatc	tttaattata	tataacttaaa	5040
cgatgtaagc	ccaaaacaaa	cgtttttttct	atatctgtct	tttgagcaaa	ttagttatac	5100
gcaaaaccaa	accgtatttta	cataaatgta	tacaaaacaa	atcgtatatt	ttcattgggt	5160
tgaaataaat	acataaaaaca	a				5181

```
<210> 13
<211> 278
<212> PRT
<213> Artificial Sequence
```

```
<220>
<223> Description of Artificial Sequence; note =
        synthetic construct
```

[illegible]

```
<210> 14
<211> 837
<212> DNA
<213> Artificial Sequence
```

```
<220>
<223> Description of Artificial Sequence; note =
```

## synthetic construct

&lt;400&gt; 14

```

atgtcgaact tcagtgcctg cgcagtgtgc ggcgatcaga gctccgggaa gcactacggc      60
gtgtcctgct gcgatgggtg ctcctgcttt ttcaagcgga gcgtgcggcg cgggagcagc      120
tacgcctgca tcgctctggt cgggaactgt gtggtggaca aggcgcggcg gaactggtgt      180
ccctcctgcc gcttccagcg atgcctggcc gtgggaatga acgctgctgc ggttcaggag      240
gagcgcggtc cgcgcaacca gcaggtggct ctctaccgca ctggccggag acaagctccg      300
ccatctcagg cggcgccatc cccgacgccc cactcccagg cgctgcactt ccagatcctc      360
gcccagatcc ttgtcacgtg cctgcgccag gcgaaggcca acgagcagtt cgctctgttg      420
gatcgctgcc aacaagacgc catcttttcag gtggtgtgga gcgagatctt cgtcctgcga      480
gcgtcccact ggtctctgga catcagcgcc atgatcgacg gctgcggcga tgagcagctc      540
aaacggctca tttgcgaggc ccaccagcta agggccgacg tcctggaact caactttatg      600
gagtccctaa tcctgtgcag aaaagaattg gccatcaatg cggagtatgc cgttatcctg      660
ggaagccact ctaaagccgc cctgatctcc ttagcccgct acaccctgca gcaatccaac      720
tacctgcggt tcggacaact gtccttgggt ctgaggcagc tgtgcctgag gcgcttcgac      780
tgcgcgcttt cttgtatgtt tcgcagcgtg gtcagggaca tcttaaaaac acttttag      837

```

&lt;210&gt; 15

&lt;211&gt; 281

&lt;212&gt; PRT

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence; note =  
synthetic construct

&lt;400&gt; 15

```

Met Gly Met Arg Arg Glu Ala Val Gln Arg Gly Arg Val Pro Pro Thr
 1          5          10          15
Gln Pro Gly Leu Ala Gly Met His Gly Gln Tyr Gln Ile Ala Asn Gly
 20          25          30
Asp Pro Met Gly Ile Ala Gly Phe Asn Gly His Ser Tyr Leu Ser Ser
 35          40          45
Tyr Ile Ser Leu Leu Leu Arg Ala Glu Pro Tyr Pro Thr Ser Arg Tyr
 50          55          60
Gly Gln Cys Met Gln Pro Asn Asn Ile Met Gly Ile Asp Asn Ile Cys
 65          70          75          80
Glu Leu Ala Ala Arg Leu Leu Phe Ser Ala Val Glu Trp Ala Lys Asn
 85          90          95
Ile Pro Phe Phe Pro Glu Leu Gln Val Thr Asp Gln Val Ala Leu Leu
100          105          110
Arg Leu Val Trp Ser Glu Leu Phe Val Leu Asn Ala Ser Gln Cys Ser
115          120          125
Met Pro Leu His Val Ala Pro Leu Leu Ala Ala Ala Gly Leu His Ala
130          135          140
Ser Pro Met Ala Ala Asp Arg Val Val Ala Phe Met Asp His Ile Arg
145          150          155          160
Ile Phe Gln Glu Gln Val Glu Lys Leu Lys Ala Leu His Val Asp Ser
165          170          175
Ala Glu Tyr Ser Cys Leu Lys Ala Ile Val Leu Phe Thr Thr Asp Ala
180          185          190
Cys Gly Leu Ser Asp Val Thr His Ile Glu Ser Leu Gln Glu Lys Ser
195          200          205
Gln Cys Ala Leu Glu Glu Tyr Cys Arg Thr Gln Tyr Pro Asn Gln Pro
210          215          220
Thr Arg Phe Gly Lys Leu Leu Leu Arg Leu Pro Ser Leu Arg Thr Val
225          230          235          240
Ser Ser Gln Val Ile Glu Gln Leu Phe Phe Val Arg Leu Val Gly Lys
245          250          255

```

Thr Pro Ile Glu Thr Leu Ile Arg Asp Met Leu Leu Ser Gly Asn Ser  
                   260                                  265                                  270  
 Phe Ser Trp Pro Tyr Leu Pro Ser Met  
                   275                                  280

<210> 16  
 <211> 2866  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Description of Artificial Sequence; note =  
       synthetic construct

<400> 16

ctaaattggt	gttttcaaaa	gaaatgaatt	tctttccact	cctttcagaa	ttcaagaata	60
aatattgaag	caatatggct	tcccttggtc	aaaccgatca	atcggtgcaa	atctttcttc	120
aagcgctcgg	tgcgacgtaa	tctaacttac	tcttgccgcg	gcagcagaaa	ctgtcccata	180
gatcaacacc	atcgcaatca	atgtcaatat	tgtcgattga	agaagtgcct	caaaatgggc	240
atgagacgcg	aagctgttca	acgtggacgc	gtaccaccca	ctcagcccgg	tctggccggc	300
atgcatgggc	agtaccagat	tgccaacggg	gatcccatgg	gcattgccgg	ctttaacggg	360
cactcgtacc	tcagttccta	catctcgctc	ctgctgcggg	cggaaaccgt	tccgacttcg	420
cgatatggcc	agtgcattga	acccaacaac	attatgggca	tcgacaacat	ctgcgaactg	480
gccgcccagc	tgtctctctc	ggcggctcgag	tgggccaaga	acataccctt	cttcccggag	540
ctgcaggtga	ccgaccaggt	ggccctgctc	cggctcgtct	ggtcagagct	cttcgtccta	600
aacgccagcc	agtgtcccat	gccgctccat	gtggcgccac	tgttgccgcg	cgcgggactt	660
catgcctccc	cgatggccgc	cgatcgtgtg	gtggccttca	tggaccacat	ccgcatcttc	720
caggagcagg	tggagaagct	gaaggcgctg	catgtcgact	ccgcggagta	ctcctgcctc	780
aaggcgatcg	tgtctctcac	caccgatgcc	tgcggcctgt	ccgatgtgac	gcacattgaa	840
tccctgcaag	agaagtcgca	gtgcgccttc	gaggaatact	gccggaccga	gtatcccac	900
cagcccacga	gattcggcaa	gctgcttctc	agactgccat	cgtgctgaac	ggctctctca	960
caagtcattg	agcaattggt	ttttgtgcgt	ctagtgcgaa	aaacgccaat	tgaaacgctg	1020
atacgcgata	tgtgtctgag	cggcaacagt	ttctcctggc	cctatctgcc	ttcgatgtga	1080
cacacgatgt	ggcgccaatt	gacaacaact	tgatcatcgg	ccgcagctgt	ggcggctgca	1140
acgctcaaca	tcaattccgg	cggaggcggc	atcggcacgc	gcggcggggg	cagtggcagt	1200
ggcgggtggcg	gtagtggagg	cgggtggcga	gtcgttggat	gtggcagcca	caacgttgct	1260
gctgccagtc	atgaccagct	cgccaatggt	gctgtcatgc	agcaaacata	cggcagcggc	1320
ggcagcagca	gcagcagcat	cagcggttgc	cacaacggta	acaacggcag	cggcggcagc	1380
atthtgaatc	agcagatcaa	caactacggc	aacaacagca	acaacaatgt	cggcaatcat	1440
atgagtgcag	gcagtttttt	cgggtgggtcc	aacaacagca	tccacagtag	tggcaatagc	1500
aataccgatt	atatgaccac	gccagccacc	gcttatgcga	caccagcgac	agcagccaca	1560
tccacgggtga	acaccacaac	gatgctgtct	aattactgcg	atgccgccac	catgatgatg	1620
gccgctgctg	cagtcaatgc	aaatcaatgc	ctgcagcaac	atcaccagcg	catgttgctc	1680
gcgggcagca	gcaacagcag	cagcaacaac	agcagcagca	acagcaacgg	cgcagcagca	1740
atgccctcct	catcctcgtc	tggctcactg	tcatctgcct	catcgacccc	aacagcaaca	1800
gcaactgcga	ctgcaattgc	aacagcaaca	gcaactgcag	cagcaacagc	cgcgcagcaa	1860
caacagcaac	aatcgccgcc	aaattttaatc	gatatcagcg	aagttcctct	cattgtggat	1920
gtcaagtagt	gtaattatth	atgcatctag	aaatggggct	ataaaccaac	cttgtagata	1980
ccccgccccg	ccccaccac	taccacaaaa	accataaaac	ccccaaaaaa	aaacaattga	2040
aaaatgtaaa	aaaaaaaaag	tggaggatga	gcgcgcgcta	gcttaattga	ctaatttttc	2100
atthttagct	tttgttgtaa	ctttgtacat	aactcctcga	aaaattcaag	tttttctcta	2160
ggccacccca	gctgtgagca	aaaccaatct	cagctgacat	atccaagaga	acttcaaaag	2220
tgaagcccc	aaaaaaagta	agaaggcgcc	aaaaaaacgt	ctttacatat	gaatgtgtat	2280
aataatttaa	tggcactgag	ttctacttaa	tttttagacca	caaacacttg	aaaaaatcaa	2340
tgaaaaaata	agaattgtgg	aaagagaaaa	atcccccccta	acactttcaa	aagacaaaac	2400
ataaagatag	ttaaaatatt	tatatatgta	atgtagcata	tacacgtata	tagtacatat	2460
atgaatatat	aaacgaaact	ctactcccag	tggtttgcag	aaatatacca	aaaattttta	2520
gctatgttta	cttgatgtgt	ggcaattttt	atgtgtgctt	tagcaatttt	atthtttact	2580
taagtaaaat	ttaaaattta	taaacattcg	attctcgact	ggtttttctc	ggcggatgta	2640
tctcaaagat	gcttctgtat	gggaaggccg	aattgttgaa	atacgaatgc	aaaattttag	2700

```

gaatttttta ttttagtaacc attacgagta aaaacacaaa atgttcagtg caagtttcag 2760
ttcttaaacg attttttcgt aagcttaagc attatcttat ttatgtgtat agagtatgaa 2820
aagttttcta tattttgtaa taataaaaat ttgcgtttat aatgaa 2866

```

&lt;210&gt; 17

&lt;211&gt; 452

&lt;212&gt; PRT

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence; note =  
synthetic construct

&lt;400&gt; 17

```

Met Gln Ser Ser Glu Gly Ser Pro Asp Met Met Asp Gln Lys Tyr Asn
 1          5          10          15
Ser Val Arg Leu Ser Pro Ala Ala Ser Ser Arg Ile Leu Tyr His Val
          20          25          30
Pro Cys Lys Val Cys Arg Asp His Ser Ser Gly Lys His Tyr Gly Ile
          35          40          45
Tyr Ala Cys Asp Gly Cys Ala Gly Phe Phe Lys Arg Ser Ile Arg Arg
          50          55          60
Ser Arg Gln Tyr Val Cys Lys Ser Gln Lys Gln Gly Leu Cys Val Val
65          70          75          80
Asp Lys Thr His Arg Asn Gln Cys Arg Ala Cys Arg Leu Arg Lys Cys
          85          90          95
Phe Glu Val Gly Met Asn Lys Asp Ala Val Gln His Glu Arg Gly Pro
          100          105          110
Arg Asn Ser Thr Leu Arg Arg His Met Ala Met Tyr Lys Asp Ala Met
          115          120          125
Met Gly Ala Gly Glu Met Pro Gln Ile Pro Ala Glu Ile Leu Met Asn
          130          135          140
Thr Ala Ala Leu Thr Gly Phe Pro Gly Val Pro Met Pro Met Pro Gly
145          150          155          160
Leu Pro Gln Arg Ala Gly His His Pro Ala His Met Ala Ala Phe Gln
          165          170          175
Pro Pro Pro Ser Ala Ala Ala Val Leu Asp Leu Ser Val Pro Arg Val
          180          185          190
Pro His His Pro Val His Gln Gly His His Gly Phe Phe Ser Pro Thr
          195          200          205
Ala Ala Tyr Met Asn Ala Leu Ala Thr Arg Ala Leu Pro Pro Thr Pro
          210          215          220
Pro Leu Met Ala Ala Glu His Ile Lys Glu Thr Ala Ala Glu His Leu
225          230          235          240
Phe Lys Asn Val Asn Trp Ile Lys Ser Val Arg Ala Phe Thr Glu Leu
          245          250          255
Pro Met Pro Asp Gln Leu Leu Leu Leu Glu Glu Ser Trp Lys Glu Phe
          260          265          270
Phe Ile Leu Ala Met Ala Gln Tyr Leu Met Pro Met Asn Phe Ala Gln
          275          280          285
Leu Leu Phe Val Tyr Glu Ser Glu Asn Ala Asn Arg Glu Ile Met Gly
          290          295          300
Met Val Thr Arg Glu Val His Ala Phe Gln Glu Val Leu Asn Gln Leu
305          310          315          320
Cys His Leu Asn Ile Asp Ser Thr Glu Tyr Glu Cys Leu Arg Ala Ile
          325          330          335
Ser Leu Phe Arg Lys Ser Pro Pro Ser Ala Ser Ser Thr Glu Asp Leu
          340          345          350
Ala Asn Ser Ser Ile Leu Thr Gly Ser Gly Ser Pro Asn Ser Ser Ala
          355          360          365

```

```

Ser Ala Glu Ser Arg Gly Leu Leu Glu Ser Gly Lys Val Ala Ala Met
   370                               380
His Asn Asp Ala Arg Ser Ala Leu His Asn Tyr Ile Gln Arg Thr His
385                               390                               395                               400
Pro Ser Gln Pro Met Arg Phe Gln Thr Leu Leu Gly Val Val Gln Leu
                               405                               410                               415
Met His Lys Val Ser Ser Phe Thr Ile Glu Glu Leu Phe Phe Arg Lys
                               420                               425                               430
Thr Ile Gly Asp Ile Thr Ile Val Arg Leu Ile Ser Asp Met Tyr Ser
                               435                               440                               445
Gln Arg Lys Ile
                               450

```

&lt;210&gt; 18

&lt;211&gt; 1885

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence; note =  
synthetic construct

&lt;400&gt; 18

```

gagtcacacat cggagtaacc aaggatatat cgaatatatc acacaatccg caataaccgcc      60
gtccacccaa accgttaaaa caaaaatcca aaacgactca aagatacacc agtgccaagt      120
gaaattcaat ttgtgcaagc gtttctacaa aaatcgccaa aattacgccc cacatcggta      180
tgcagtcgtc ggagggttca ccagacatga tggatcagaa atacaacagc gtgcgtcttt      240
cgccagcggc atcgagtcgc attctatacc atgtgccctg caaagtctgc agagatcaca      300
gctccggcaa gcattacggc atctacgcct gtgatggctg cgccggattc ttcaagagga      360
gcattcggag atcccggcag tatgtgtgca agtcgcagaa gcagggactc tgtgtggtgg      420
acaagacgea caggaaccaa tgtagggtct gccgactgag gaagtgcctt gaggtcggaa      480
tgaacaagga tgcagtgcag cagcagcggg gaccgcggaa ctccactctg cgtcgccaca      540
tggccatgta caaggatgcc atgatgggag ccggcgagat gccacaaata ccgcccgaag      600
ttctgatgaa cacggctgcc ttgaccggct ttcttgagat accgatgccc atgcctggcc      660
tgccccagag ggctggatcat catcctgctc acatggctgc cttccagccg ccaccatcgg      720
ctgccgctgt cttggactta tccgtgccac gagtgcccca tcaccgggtg caccaaggac      780
accacggttt cttctcgccc accgccgcct acatgaatgc cctggccact cgggccctgc      840
ccccactcc tccgctgatg gcagctgagc acatcaagga aaccgcggcg gaacacctat      900
tcaagaacgt caactggatc aagagcgtac gggccttcac cgaactgccc atgccggatc      960
agctgctcct gctggaggag tcctggaagg agttcttcat cctggccatg gccagtagc      1020
taatgcccac gaatttcgcc cagctgctgt tcgtctacga gtccgagaat gccaacggg      1080
agatcatggg catggtgacc cgcgaggtgc acgccttcca ggagggtgctg aaccaactgt      1140
gccatctgaa cattgacagc accgagtagc agtgtctgag ggctatttcg ctcttccgta      1200
agtcaccacc gtcggcaagt tctaccgagg atttagccaa cagctcaatc ctgacaggaa      1260
gcggcagccc gaactcctcg gcctctgctg aatccagggg tcttctggag tcgggaaaag      1320
tgggcgccat gcacaacgat gcccgaggtg cgctgcacaa ctacatccag aggacccatc      1380
cctcgcagcc catgcgattc cagacgctct tgggcgtggt gcagctgatg cacaaggctc      1440
caagcttcac catcgaggag ctgttcttcc gaaagacatc cggcgacatc accattgtgc      1500
gcctcatctc cgacatgtac agtcagcgca agatctgaaa agtatgtaga gcctagacta      1560
atcgccgcac tcgaagtgcc ttccaagtgc tgggaactgt gataatctcg gaagaagcgc      1620
tttggacaat actcgatcag tgaaatcaac gatttctcat atccaggagt cgagccttaa      1680
aatacgtaca caacactcac ctttaatacct tacctaaaca gaactcgaag taatcttagc      1740
taaagtctct cagaccatcc agatgtgttt caaattgcat tcgcaaaagt ttcaactttg      1800
cctgttaaat acgtcaatcg tagttttaa cacttttagt ttaagcgcac attattagct      1860
ttaggatttg gaaaaataat tattc                                     1885

```

&lt;210&gt; 19

&lt;211&gt; 691

&lt;212&gt; PRT

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence; note =  
synthetic construct.

&lt;400&gt; 19

Met	Gly	Thr	Ala	Gly	Asp	Arg	Leu	Leu	Asp	Ile	Pro	Cys	Lys	Val	Cys
1				5					10					15	
Gly	Asp	Arg	Ser	Ser	Gly	Lys	His	Tyr	Gly	Ile	Tyr	Ser	Cys	Asp	Gly
			20					25					30		
Cys	Ser	Gly	Phe	Phe	Lys	Arg	Ser	Ile	His	Arg	Asn	Arg	Ile	Tyr	Thr
		35					40					45			
Cys	Lys	Ala	Thr	Gly	Asp	Leu	Lys	Gly	Arg	Cys	Pro	Val	Asp	Lys	Thr
	50					55					60				
His	Arg	Asn	Gln	Cys	Arg	Ala	Cys	Arg	Leu	Ala	Lys	Cys	Phe	Gln	Ser
65					70					75					80
Ala	Met	Asn	Lys	Asp	Ala	Val	Gln	His	Glu	Arg	Gly	Pro	Arg	Lys	Pro
				85					90					95	
Lys	Leu	His	Pro	Gln	Leu	His	His	His	His	His	His	Ala	Ala	Ala	Ala
			100					105					110		
Ala	Ala	Ala	Ala	His	His	Ala	Ala	Ala	Ala	His	His	His	His	His	His
			115					120					125		
His	His	His	Ala	His	Ala	Ala	Ala	Ala	His	His	Ala	Ala	Val	Ala	Ala
			130				135						140		
Ala	Ala	Ala	Ser	Gly	Leu	His	His	His	His	His	Ala	Met	Pro	Val	Ser
145					150					155					160
Leu	Val	Thr	Asn	Val	Ser	Ala	Ser	Phe	Asn	Tyr	Thr	Gln	His	Ile	Ser
			165						170					175	
Thr	His	Pro	Pro	Ala	Pro	Ala	Ala	Pro	Pro	Ser	Gly	Phe	His	Leu	Thr
			180					185					190		
Ala	Ser	Gly	Ala	Gln	Gln	Gly	Pro	Ala	Pro	Pro	Ala	Gly	His	Leu	His
			195				200					205			
His	Gly	Gly	Ala	Gly	His	Gln	His	Ala	Thr	Ala	Phe	His	His	Pro	Gly
	210					215					220				
His	Gly	His	Ala	Leu	Pro	Ala	Pro	His	Gly	Gly	Val	Val	Ser	Asn	Pro
225					230					235					240
Gly	Gly	Asn	Ser	Ser	Ala	Ile	Ser	Gly	Ser	Gly	Pro	Gly	Ser	Thr	Leu
			245						250					255	
Pro	Phe	Pro	Ser	His	Leu	Leu	His	His	Asn	Leu	Ile	Ala	Glu	Ala	Ala
			260					265					270		
Ser	Lys	Leu	Pro	Gly	Ile	Thr	Ala	Thr	Ala	Val	Ala	Ala	Val	Val	Ser
			275				280					285			
Ser	Thr	Ser	Thr	Pro	Tyr	Ala	Ser	Ala	Ala	Gln	Thr	Ser	Ser	Pro	Ser
	290					295					300				
Ser	Asn	Asn	His	Asn	Tyr	Ser	Ser	Pro	Ser	Pro	Ser	Asn	Ser	Ile	Gln
305					310					315					320
Ser	Ile	Ser	Ser	Ile	Gly	Ser	Arg	Ser	Gly	Gly	Gly	Glu	Glu	Gly	Leu
			325						330					335	
Ser	Leu	Gly	Ser	Glu	Ser	Pro	Arg	Val	Asn	Val	Glu	Thr	Glu	Thr	Pro
			340					345					350		
Ser	Pro	Ser	Asn	Ser	Pro	Pro	Leu	Ser	Ala	Gly	Ser	Ile	Ser	Pro	Ala
			355				360					365			
Pro	Thr	Leu	Thr	Thr	Ser	Ser	Gly	Ser	Pro	Gln	His	Arg	Gln	Met	Ser
	370					375					380				
Arg	His	Ser	Leu	Ser	Glu	Ala	Thr	Thr	Pro	Pro	Ser	His	Ala	Ser	Leu
385					390					395					400
Met	Ile	Cys	Ala	Ser	Asn	Asn	Asn	Asn	Asn	Asn	Asn	Asn	Asn	Asn	Asn
			405						410					415	
Asn	Gly	Glu	His	Lys	Gln	Ser	Ser	Tyr	Thr	Ser	Gly	Ser	Pro	Thr	Pro
			420					425					430		
Thr	Thr	Pro	Thr	Pro	Pro	Pro	Pro	Arg	Ser	Gly	Val	Gly	Ser	Thr	Cys
		435					440					445			

Asn Thr Ala Ser Ser Ser Ser Gly Phe Leu Glu Leu Leu Leu Ser Pro  
 450 455 460  
 Asp Lys Cys Gln Glu Leu Ile Gln Tyr Gln Val Gln His Asn Thr Leu  
 465 470 475 480  
 Leu Phe Pro Gln Gln Leu Leu Asp Ser Arg Leu Leu Ser Trp Glu Met  
 485 490 495  
 Leu Gln Glu Thr Thr Ala Arg Leu Leu Phe Met Ala Val Arg Trp Val  
 500 505 510  
 Lys Cys Leu Met Pro Phe Gln Thr Leu Ser Lys Asn Asp Gln His Leu  
 515 520 525  
 Leu Leu Gln Glu Ser Trp Lys Glu Leu Phe Leu Leu Asn Leu Ala Gln  
 530 535 540  
 Trp Thr Ile Pro Leu Asp Leu Thr Pro Ile Leu Glu Ser Pro Leu Ile  
 545 550 555 560  
 Arg Glu Arg Val Leu Gln Asp Glu Ala Thr Gln Thr Glu Met Lys Thr  
 565 570 575  
 Ile Gln Glu Ile Leu Cys Arg Phe Arg Gln Ile Thr Pro Asp Gly Ser  
 580 585 590  
 Glu Val Gly Cys Met Lys Ala Ile Ala Leu Phe Ala Pro Glu Thr Ala  
 595 600 605  
 Gly Leu Cys Asp Val Gln Pro Val Glu Met Leu Gln Asp Gln Ala Gln  
 610 615 620  
 Cys Ile Leu Ser Asp His Val Arg Leu Arg Tyr Pro Arg Gln Ala Thr  
 625 630 635 640  
 Arg Phe Gly Arg Leu Leu Leu Leu Leu Pro Ser Leu Arg Thr Ile Arg  
 645 650 655  
 Ala Ala Thr Ile Glu Ala Leu Phe Phe Lys Glu Thr Ile Gly Asn Val  
 660 665 670  
 Pro Ile Ala Arg Leu Leu Arg Asp Met Tyr Thr Met Glu Pro Ala Gln  
 675 680 685  
 Val Asp Lys  
 690

&lt;210&gt; 20

&lt;211&gt; 3043

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

 <223> Description of Artificial Sequence; note =  
 synthetic construct

&lt;400&gt; 20

```

gtcagcccag gcgatccgca tttgcgtccg cagcagggtt ccgatttcag aactctgatt      60
ccagcggcag cgaatcgcggt cggcatctga acatttgaaa ataatctaaa attgcaagtg      120
actttgtgca ccggttacac taaaattggt aacaaatcgc catatattct gaatttaaatt      180
ttaaagtgcg cagtgcggaa tataaatcag agcaaactgg atacgttagg gttcaaatac      240
ttccatcaac ggaaaatggg cacagcgggc gatcgcctgt tggacattcc ctgcaagggtg      300
tgtggcgatc gcagctccgg caagcactat ggaatctaca gctgcgatgg ctgctccggt      360
tttttcaagc ggagcattca tcgcaatcgg atttacacct gtaaggccac cggcgatctc      420
aagggtcgct gtccggtgga caagacccat cggaatcagt gtcgcgcctg tcgcctggcc      480
aagtgtctcc agtcggccat gaacaaggat gctgtgcagc acgagcgcgg tcctaggaaa      540
ccaagtgtgc acccgcaact gcatcatcat catcatcatg ctgctgccgc cgccgctgca      600
gcgcatcatg cagcagccgc ccatcaccat caccatcatc accaccacgc ccacgcagcg      660
gccgcccata atgcggcagt ggctgcagcg gctgcctccg ggctgcatca ccaccaccac      720
gccatgcccg tctcgtggtt gaccaatgtc tcggcctcgt tcaactatac gcagcacatc      780
tccacgcata cgcctgctcc ggcggcgcca cccagtggct ttcacctgac ggccagtggc      840
gccagcagg gaccagctcc accagctggc cacctgcacc atggtggagc cggacatcag      900
cacgccacgg ccttccacca tccgggacat ggacacgcgc tgcttgcgcc acatggcggc      960
gtcgtcagca atcccggcgg caactcgagc gcaatctccg gcagcgggtc cggctccacg     1020
ctgcccttcc cctcgcacct gctgcaccac aatctgatag cggaggcggc cagcaagctg     1080

```

```

ccgggcatca ctgccacagc cgttgcggcg gtggtgtcct ccactagcac gccctacgcc 1140
tcggcgggccc agacgtcgtc gcctagtagc aacaaccaca actactcctc gccctcgccc 1200
agcaactcca tccagtcctat ctcgagcatt ggatcgcgca gcggtggtgg cgaggagggc 1260
ctcagcctgg gcagcgagag tccgcgcgtc aatgtggaaa cggagacacc ttcgccatcg 1320
aactcgccgc cccttagtgc tggtagcatt tcgccagcgc ccacgttgac cacctcgtcg 1380
ggatcgccgc agcaccgcca gatgtcgcgg cacagcctca gtgaggcaac cagcccgccc 1440
agccacgcct ctctcatgat ttgcgccagc aacaataaca ataacaacaa taataataac 1500
aataatggag agcacaagca gtcgagctac acatccggat caccgacacc cacaacgccc 1560
acgccgcccac cgccgcgttc tgggtgtagg tccacctgca acacggccag cagctccagc 1620
ggcttcctgg agctgctgct cagtcgggac aagtgccagg agctcatcca gtaccagggtg 1680
cagcacaaca cgctgctctt cccgcaacag ctglttggaet cgcggctgct ctcttgggag 1740
atgctgcagg agacgacggc gcgactgctc ttcattggcg tgcgctgggt caagtgcctc 1800
atgcccttcc agacgctctc caagaacgac cagcatttgc tgctccagga atcctggaag 1860
gagctcttcc tgctcaacct cgcccaatgg actataccgc tggatctaac gcccatactg 1920
gaatcacccg tcattccgca acgggtgctg caggacgagg ccacacaaac ggagatgaag 1980
acgatccagg agatcctctg ccgcttccgc cagatcacac ccgacggcag cgagggtgggc 2040
tgcatgaagg ccacgcctt gtctgcaccc gaaaccgccc gcctgtgcga cgtgcagccg 2100
gtggagatgt tgcaggatca ggcgcagtgc atcctctccg accatgtgct actgcgctac 2160
cctcgccaag caaccgctt cggcaggctg ctgctcctgc tgccctcgct gcgcaccatc 2220
cgggcggcca ccacgagggc gctgttcttc aaggagacca tcggcaatgt gccattgct 2280
cgactgctgc gcgacatgta caccatggaa ccggcacagg tggacaagtg aaccggccac 2340
gcatgacagt cgaaatgaaa tcaaaatcga ttccctagca cctaagcgcc acccatcggt 2400
cgctgctcata tgcgaactta tttgtattcc aatgcgaccc gaatcctatt cagattcact 2460
gcggcaggag gcggtccaaa tgtggggcgg aagctgcaga tgctatgggt cgcaggacgc 2520
catgtaatgg aggcgtatgt actaacgcgc ctctccatt ggcgatgcag tccgcgatga 2580
tggcgcactc ccacaccac acccgtaccc acaccttgat ttatcgccgg caatgcgtcg 2640
gagtcctcct actttcgtt cgttttctaa catttgtatc cttattttat ttcattttt 2700
tccacggatt tttcgttttg actgcctggg cggcactcct tatttatctt tcattcgacg 2760
ttttgtcgtc gcttttctaa aaattcccca tgttatttca acctggcaag gacctcgcag 2820
tccatttccc gcgcccttac ttacaaatca cttcccatcc cacatccagc aattccgtgg 2880
tttgaattct ttctgtcatt gactacgaaa taccctttta tcagacaaat aaagaatatt 2940
agttgtaatt cttttttctg caatccagct ctaaaacggg tttcttaatc gaaatcgata 3000
aatgtaaaaa ttatacatat cctttaccaa cattgtttgc cta 3043

```

&lt;210&gt; 21

&lt;211&gt; 532

&lt;212&gt; PRT

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence; note =  
synthetic construct

&lt;400&gt; 21

```

Met Ala Thr Gly Arg Ser Leu Leu Phe Arg Val Pro Trp Tyr Val Cys
 1             5             10             15
Leu Cys Val Cys Ala Glu Ser Ala Glu Pro Gly Val Tyr Trp Arg Leu
          20             25             30
Arg Leu Arg Leu Gly Leu Pro Thr Leu Ala Gly Pro His Thr Asn Thr
          35             40             45
Leu Thr Leu Thr Ala Arg Thr Ser Ser Cys Arg Ser Ile Lys Lys Glu
          50             55             60
Arg Ile Lys Ala Ser Gln Gln Ala Asn Ala Pro Pro Glu Leu Pro Leu
65             70             75             80
Lys Val Ser Val Asp Val Asn Ile Ile Ile Ala Ala His Ser Gln Arg
          85             90             95
Arg Arg Ile Gly Leu Val Arg Phe His Gln Arg Glu Ser Glu Asp Arg
          100            105            110
Pro Leu Ala Val Ala Ser Pro Arg Leu Gln Ile Asn Met Glu Pro Thr
          115            120            125

```

Ala	Met	Asn	Pro	Lys	Lys	Leu	His	Ser	Pro	Gln	Arg	His	Cys	Tyr	Thr
130						135					140				
Pro	Pro	Pro	Ala	Pro	Met	His	Gly	Gln	Ala	Pro	Pro	Pro	Thr	Ser	Thr
145					150					155					160
Gly	Val	Ala	Pro	Pro	Thr	Gln	Pro	Pro	Pro	Pro	His	Pro	Ala	Ala	Pro
				165						170					175
Asn	Val	Pro	Asn	Gly	Arg	Leu	Leu	Ser	Trp	Asn	His	Ser	Ala	Ala	Ala
			180					185					190		
Ala	Ala	Ala	Ala	Ala	Ala	Ala	Gln	Ala	Ala	Ala	Asn	Ser	Met	Asn	His
		195					200					205			
Ser	Ser	Ala	Ala	Glu	Gly	Ser	Ser	Met	Thr	Arg	Ile	Lys	Gly	Gln	Asn
210					215						220				
Leu	Gly	Leu	Ile	Cys	Val	Val	Cys	Gly	Asp	Thr	Ser	Ser	Gly	Lys	His
225				230					235						240
Tyr	Gly	Ile	Leu	Ala	Cys	Asn	Gly	Cys	Ser	Gly	Phe	Phe	Lys	Arg	Ser
				245					250					255	
Val	Arg	Arg	Lys	Leu	Ile	Tyr	Arg	Cys	Gln	Ala	Gly	Thr	Gly	Arg	Cys
			260					265					270		
Val	Val	Asp	Lys	Ala	His	Arg	Asn	Gln	Cys	Gln	Ala	Cys	Arg	Leu	Lys
		275					280					285			
Lys	Cys	Leu	Gln	Met	Gly	Met	Asn	Lys	Asp	Asp	Asp	Ser	Ile	Asp	Val
290				295							300				
Thr	Asn	Asp	Asn	Glu	Glu	Pro	His	Ala	Val	Ser	Arg	Ser	Asp	Ser	Ser
305				310						315					320
Phe	Ile	Met	Pro	Gln	Phe	Met	Ser	Pro	Asn	Leu	Tyr	Thr	His	Gln	His
				325					330					335	
Glu	Thr	Val	Tyr	Glu	Thr	Ser	Ala	Arg	Leu	Leu	Phe	Met	Ala	Val	Lys
			340					345					350		
Trp	Ala	Lys	Asn	Leu	Pro	Ser	Phe	Ala	Arg	Leu	Ser	Phe	Arg	Asp	Gln
		355					360					365			
Val	Ile	Leu	Leu	Glu	Glu	Ser	Trp	Ser	Glu	Leu	Phe	Leu	Leu	Asn	Ala
370				375							380				
Ile	Gln	Trp	Cys	Ile	Pro	Leu	Asp	Pro	Thr	Gly	Cys	Ala	Leu	Phe	Ser
385				390						395					400
Val	Ala	Glu	His	Cys	Asn	Asn	Leu	Glu	Asn	Asn	Ala	Asn	Gly	Asp	Thr
				405					410					415	
Cys	Ile	Thr	Lys	Glu	Glu	Leu	Ala	Ala	Asp	Val	Arg	Thr	Leu	His	Glu
			420					425					430		
Ile	Phe	Cys	Lys	Tyr	Lys	Ala	Val	Leu	Val	Asp	Pro	Ala	Glu	Phe	Ala
		435					440					445			
Cys	Leu	Lys	Ala	Ile	Val	Leu	Phe	Arg	Pro	Glu	Thr	Arg	Gly	Leu	Lys
450				455						460					
Asp	Pro	Ala	Gln	Ile	Glu	Asn	Leu	Gln	Asp	Gln	Ala	His	His	Thr	Lys
465				470						475					480
Thr	Gln	Phe	Thr	Ala	Gln	Ile	Ala	Arg	Phe	Gly	Arg	Leu	Leu	Leu	Met
				485					490					495	
Leu	Pro	Leu	Leu	Arg	Met	Ile	Ser	Ser	His	Lys	Ile	Glu	Ser	Ile	Tyr
			500					505					510		
Phe	Gln	Arg	Thr	Ile	Gly	Asn	Thr	Pro	Met	Glu	Lys	Val	Leu	Cys	Asp
		515					520					525			
Met	Tyr	Lys	Asn												
530															

&lt;210&gt; 22

&lt;211&gt; 1599

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence; note =  
synthetic construct

&lt;400&gt; 22

```

atggcgaccg ggcgttctct gctctttcga gtgccttggt atgtgtgctt gtgtgtgtgc      60
gcagagagcg cagagccggg tgtttatttg agattgcgat tgcggcttgg cttacccaca      120
ctcgcagggc cgcacaccaa cacactaaca ctaacagcga ggacaagctc ctgccgcagc      180
atcaagaagg aacgaatcaa agcaagccaa caagcaaatg cgcaccaga gttgccacta      240
aaagtctccg ttgacgttaa catcatcatc gcggcacact cgcagcgccg tcggatcgga      300
ttggttcggg ttcatcagcg ggaatcagag gaccgtccac ttgccgtcgc ctctccacga      360
ttgcaaatta atatggagcc tactgcgatg aaccgaaaaa aactccacag tccgcagcgg      420
cattgctaca ctccgccgcc ggccgccgatg cacggacagg cgcctccacc tacatcaacg      480
ggcgtggccc cgcccacaca gccaccgccc cctcatcccg ccgccccaaa cgtgcccaat      540
ggtcgattgc tgagctggaa tcacagtgcc gctgcagctg ctgcggcggc ggacgcccac      600
gcggcagcca actccatgaa ccactcgtcg gcggcggagg gttcatcgat gaccgaggatt      660
aagggtcaga acctgggcct catctgcgtg gtgtgcggcg acaccagctc gggaaagcac      720
tacggaatcc tagcctgcaa tggctgctcc ggattcttca aacgcagcgt gcggcggaaa      780
ctcatttatc gctgccaggc gggaacggga cgctgtgtgg tggacaaagc tcatcggaat      840
caatgccagg cctgcaggct caagaagtgc cttcaaatgg gaatgaacaa ggacgacgac      900
tccatagatg taaccaacga caacgaggag ccgcatgcag tcagcagatc ggattcgagt      960
ttcattatgc cgcagttcat gtcgcccaat ctgtacaccc atcaacacga aacagtttac     1020
gagacaagtg cccggctgct cttcatggcc gtcaagtggg ccaagaacct gccagctttt     1080
gcaagacttt cctttcggga tcaggtaatt ttgctggagg agtcctggtc ggagctgttc     1140
ctgctgaacg caatccaatg gtgcattccc ctggatccca ccggctgcgc cctcttctcg     1200
gtggcggagc actgcaataa tctagagaac aatgccaatg gcgacacttg cataacaaag     1260
gaggagctgg cggcggatgt gcgaacgctc cacgagatct tctgcaaata caaggcggtg     1320
ctggtggacc ccgctgaatt cgcgtgcctc aaggcgatag ttctcttccg gccggaaacg     1380
cgcgactta aagatccggc gcagatagag aatcttcagg atcaggcgca ccacacaaag     1440
acgcagtcca ccgccagat agccagattc ggacgactcc ttctcatgct gccgttgctg     1500
cgcatgatca gctcccacaa gattgagtc atctattttc agcgcactat tgggaacacg     1560
cccatggaaa aggtgctctg tgacatgtat aagaactag                                1599

```

&lt;210&gt; 23

&lt;211&gt; 484

&lt;212&gt; PRT

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence; note =  
synthetic construct

&lt;400&gt; 23

```

Met Ser Asp Gly Val Ser Ile Leu His Ile Lys Gln Glu Val Asp Thr
 1              5              10              15
Pro Ser Ala Ser Cys Phe Ser Pro Ser Ser Lys Ser Thr Ala Thr Gln
              20              25              30
Ser Gly Thr Asn Gly Leu Lys Ser Ser Pro Ser Val Ser Pro Glu Arg
              35              40              45
Gln Leu Cys Ser Ser Thr Thr Ser Leu Ser Cys Asp Leu His Asn Val
              50              55              60
Ser Leu Ser Asn Asp Gly Asp Ser Leu Lys Gly Ser Gly Thr Ser Gly
65              70              75              80
Gly Asn Gly Gly Gly Gly Gly Gly Gly Thr Ser Gly Gly Asn Ala Thr
              85              90              95
Asn Ala Ser Ala Gly Ala Gly Ser Gly Ser Val Arg Asp Glu Leu Arg
              100             105             110
Arg Leu Cys Leu Val Cys Gly Asp Val Ala Ser Gly Phe His Tyr Gly
              115             120             125
Val Ala Ser Cys Glu Ala Cys Lys Ala Phe Phe Lys Arg Thr Ile Gln
              130             135             140
Gly Asn Ile Glu Tyr Thr Cys Pro Ala Asn Asn Glu Cys Glu Ile Asn
145             150             155             160

```

Lys	Arg	Arg	Arg	Lys	Ala	Cys	Gln	Ala	Cys	Arg	Phe	Gln	Lys	Cys	Leu	
				165					170					175		
Leu	Met	Gly	Met	Leu	Lys	Glu	Gly	Val	Arg	Leu	Asp	Arg	Val	Arg	Gly	
			180					185					190			
Gly	Arg	Gln	Lys	Tyr	Arg	Arg	Asn	Pro	Val	Ser	Asn	Ser	Tyr	Gln	Thr	
		195					200					205				
Met	Gln	Leu	Leu	Tyr	Gln	Ser	Asn	Thr	Thr	Ser	Leu	Cys	Asp	Val	Lys	
	210					215					220					
Ile	Leu	Glu	Val	Leu	Asn	Ser	Tyr	Glu	Pro	Asp	Ala	Leu	Ser	Val	Gln	
225					230					235					240	
Thr	Pro	Pro	Pro	Gln	Val	His	Thr	Thr	Ser	Ile	Thr	Asn	Asp	Glu	Ala	
				245					250					255		
Ser	Ser	Ser	Ser	Gly	Ser	Ile	Lys	Leu	Glu	Ser	Ser	Val	Val	Thr	Pro	
			260					265					270			
Asn	Gly	Thr	Cys	Ile	Phe	Gln	Asn	Asn	Asn	Asn	Asn	Asp	Pro	Asn	Glu	
	275						280					285				
Ile	Leu	Ser	Val	Leu	Ser	Asp	Ile	Tyr	Asp	Lys	Glu	Leu	Val	Ser	Val	
	290					295					300					
Ile	Gly	Trp	Ala	Lys	Gln	Ile	Pro	Gly	Phe	Ile	Asp	Leu	Pro	Leu	Asn	
305					310					315					320	
Asp	Gln	Met	Lys	Leu	Leu	Gln	Val	Ser	Trp	Ala	Glu	Ile	Leu	Thr	Leu	
				325					330					335		
Gln	Leu	Thr	Phe	Arg	Ser	Leu	Pro	Phe	Asn	Gly	Lys	Leu	Cys	Phe	Ala	
			340					345					350			
Thr	Asp	Val	Trp	Met	Asp	Glu	His	Leu	Ala	Lys	Glu	Cys	Gly	Tyr	Thr	
	355					360						365				
Glu	Phe	Tyr	Tyr	His	Cys	Val	Gln	Ile	Ala	Gln	Arg	Met	Glu	Arg	Ile	
	370					375					380					
Ser	Pro	Arg	Arg	Glu	Glu	Tyr	Tyr	Leu	Leu	Lys	Ala	Leu	Leu	Leu	Ala	
385					390					395					400	
Asn	Cys	Asp	Ile	Leu	Leu	Asp	Asp	Gln	Ser	Ser	Leu	Arg	Ala	Phe	Arg	
			405					410					415			
Asp	Thr	Ile	Leu	Asn	Ser	Leu	Asn	Asp	Val	Val	Tyr	Leu	Leu	Arg	His	
			420					425					430			
Ser	Ser	Ala	Val	Ser	His	Gln	Gln	Gln	Leu	Leu	Leu	Leu	Leu	Pro	Ser	
		435					440					445				
Leu	Arg	Gln	Ala	Asp	Asp	Ile	Leu	Arg	Arg	Phe	Trp	Arg	Gly	Ile	Ala	
	450					455					460					
Arg	Asp	Glu	Val	Ile	Thr	Met	Lys	Lys	Leu	Phe	Leu	Glu	Met	Leu	Glu	
465					470					475					480	
Pro	Leu	Ala	Arg													

&lt;210&gt; 24

&lt;211&gt; 2529

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence; note =  
synthetic construct

&lt;400&gt; 24

ccctgggtcag	gtctgggttca	ccaaaaaaga	aaataaaatt	acattttcaat	ctttccaata	60
tgcaaataatc	tgcacgaaaa	ccagcgagaa	cagcatgctc	acaataaaga	gcccccaaac	120
aatgtgactc	gtatccgcgc	agagtgacgt	ttcgtgcctt	gcccgagtgc	caaatccaaa	180
tccaatcca	ggcgcacaaa	atcgatgcag	atgctgtctg	catttctcata	gaaagtgcga	240
ctgaataacc	gatggtcgcc	aaaagccacg	atgtccagta	ataatgacca	gtgaataaac	300
aattatgact	cgagcatcga	aaaatgctga	ggaacgaata	cataagcaat	aacaagaagg	360
tgctcaactc	ggaccaaaaac	aagtactaca	tgctaacggt	cgaggaggcc	gatatgtatt	420
gacgttggtta	cagtggagct	gattacacaa	aagatcctca	gaacgatttt	atccaaggca	480

```

cgaacatgtc cgacggcgtc agcatcttgc acatcaaaca ggaggtggac actccatcgg 540
cgtcctgctt tagtcccagc tccaagtcaa cggccacgca gagtggcaca aacggcctga 600
aatcctcgcc ctcggtttcg ccggaaaggc agctctgcag ctcgacgacc tctctatcct 660
gcgatttgca caatgtatcc ttaagcaatg atggcgatag tctgaaagga agtggtagaa 720
gtggcgggcaa tggcgaggga ggaggtggtg gtacgagtgg tggaaatgcg accaatgcga 780
gtgccggagc tggatcggga tccgtcaggg acgagctcgg ccgatttgtg ttggtttgtg 840
gcgatgtggc cagtggattc cactatggtg tggcgagttg tgaggcttgc aaagcgttct 900
ttaaacgcac catccaaggc aacatcgagt acacgtgtcc ggccaacaac gagtgtgaga 960
ttaacaagcg gagacgcaag gcctgccaaag cgtgtcgctt ccagaaatgt ctactaatgg 1020
gcatgctcaa ggaggggtgtg cgcttgatc gagttcgtgg aggacggcag aagtaccgaa 1080
ggaatcctgt atcaaaactct taccagacta tgcagctgct ataccaatcc aacaccacct 1140
cgctgtgcga tgtcaagata ctggaggtgc tcaattcata tgagccggat gccttgagcg 1200
tccaaacgcc gccgccgcaa gtccacacga ctagcataac taatgatgag gcctcatcct 1260
cctcgggagc cataaaactg gagtccagcg ttgttacgcc caatgggact tgcattttcc 1320
aaaacaacaa caacaatgat cccaatgaga tactaagcgt ccttagtgat atttacgaca 1380
aggaattggt cagcgtcatt ggctgggcca agcagatacc tggctttata gatctgccac 1440
ttaacgacca gatgaagctt ctccaggtgt cgtgggcaga gatcctgacg ctccagctga 1500
ccttcgggtc cctaccgttc aatggcaagt tatgcttcgc cacggatgtc tggatggatg 1560
aacatttggc caaggagtgc ggttacacgg agttctacta ccactgcgtc cagatcgcac 1620
agcgcatgga aagaatatcg ccacgaaggg aggagtacta cttgctaaag gcgctcctgc 1680
tggccaactg cgacattctg ctggatgatc agagttccct gcgcgcattt cgtgatacga 1740
ttcttaattc tctaaacgat gtggtctact tgctgcgtca ttcgtcggcc gtgtcgcac 1800
agcaacaatt gctgcttttg ctgccttcgc tgcggcaggg ggatgatata ctgcgaagat 1860
tttggcgtgg aattgcacgc gatgaagtca ttaccatgaa gaaactgttc ctcgagatgc 1920
tcgagccgct ggccaggtga aaaggattat gcgggcgccc aaactagttg atctagctga 1980
taagcaaagg tgcaaatata gtcttaggta tatatggatg tatactagag tagattaagc 2040
gtaggataag ccatgtatat aaatagtaaa atacttgctg ggtaagatta gttcgcagaa 2100
aaaatctctt ttaatggact accaactaca gcaactggaa aaccctactt atcttctaga 2160
atcgggggtg gcttacactg gttaaaggcg catatagggt ttatgtgtct aaagtgtga 2220
gtcacagatc ttcaataatt tgttcaattc tcaactggtc tgatatatgt atatgccgca 2280
accttctgat gtaacgtatg aatttgtggg cactttttaa atacgatagt ggttctacaa 2340
tacaatggat tatactgttt ctaagtgtca tgtaaccagc tgattctgtg tctatgtggt 2400
acacatgcgg tcaaaagaat agcaatgtcg tccgtgaata ataaaccgtt tgtaactgtt 2460
gtttccatac tccctaagtt ctgtattctt tggggatttt cttttcctaa acaaattcaa 2520
attagtttt 2529

```

&lt;210&gt; 25

&lt;211&gt; 601

&lt;212&gt; PRT

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence; note =  
synthetic construct

&lt;400&gt; 25

```

Met Asp Gly Val Lys Val Glu Thr Phe Ile Lys Ser Glu Glu Asn Arg
 1           5           10          15
Ala Met Pro Leu Ile Gly Gly Gly Ser Ala Ser Gly Gly Thr Pro Leu
          20          25          30
Pro Gly Gly Gly Val Gly Met Gly Ala Gly Ala Ser Ala Thr Leu Ser
          35          40          45
Val Glu Leu Cys Leu Val Cys Gly Asp Arg Ala Ser Gly Arg His Tyr
          50          55          60
Gly Ala Ile Ser Cys Glu Gly Cys Lys Gly Phe Phe Lys Arg Ser Ile
65          70          75          80
Arg Lys Gln Leu Gly Tyr Gln Cys Arg Gly Ala Met Asn Cys Glu Val
          85          90          95
Thr Lys His His Arg Asn Arg Cys Gln Phe Cys Arg Leu Gln Lys Cys
          100         105         110

```

Leu	Ala	Ser	Gly	Met	Arg	Ser	Asp	Ser	Val	Gln	His	Glu	Arg	Lys	Pro
		115					120					125			
Ile	Val	Asp	Arg	Lys	Glu	Gly	Ile	Ile	Ala	Ala	Ala	Gly	Ser	Ser	Ser
	130					135					140				
Thr	Ser	Gly	Gly	Gly	Asn	Gly	Ser	Ser	Thr	Tyr	Leu	Ser	Gly	Lys	Ser
145					150					155					160
Gly	Tyr	Gln	Gln	Gly	Arg	Gly	Lys	Gly	His	Ser	Val	Lys	Ala	Glu	Ser
				165					170						175
Ala	Ala	Thr	Pro	Pro	Val	His	Ser	Ala	Pro	Ala	Thr	Ala	Phe	Asn	Leu
			180					185					190		
Asn	Glu	Asn	Ile	Phe	Pro	Met	Gly	Leu	Asn	Phe	Ala	Glu	Leu	Thr	Gln
	195						200					205			
Thr	Leu	Met	Phe	Ala	Thr	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	His
	210					215					220				
Gln	Gln	Ser	Gly	Ser	Tyr	Ser	Pro	Asp	Ile	Pro	Lys	Ala	Asp	Pro	Glu
225					230					235					240
Asp	Asp	Glu	Asp	Asp	Ser	Met	Asp	Asn	Ser	Ser	Thr	Leu	Cys	Leu	Gln
				245					250						255
Leu	Leu	Ala	Asn	Ser	Ala	Ser	Asn	Asn	Asn	Ser	Gln	His	Leu	Asn	Phe
			260					265					270		
Asn	Ala	Gly	Glu	Val	Pro	Thr	Ala	Leu	Pro	Thr	Thr	Ser	Thr	Met	Gly
		275					280					285			
Leu	Ile	Gln	Ser	Ser	Leu	Asp	Met	Arg	Val	Ile	His	Lys	Gly	Leu	Gln
	290					295					300				
Ile	Leu	Gln	Pro	Ile	Gln	Asn	Gln	Leu	Glu	Arg	Asn	Gly	Asn	Leu	Ser
305					310					315					320
Val	Lys	Pro	Glu	Cys	Asp	Ser	Glu	Ala	Glu	Asp	Ser	Gly	Thr	Glu	Asp
				325					330					335	
Ala	Val	Asp	Ala	Glu	Leu	Glu	His	Met	Glu	Leu	Asp	Phe	Glu	Cys	Gly
			340					345					350		
Gly	Asn	Arg	Ser	Gly	Gly	Ser	Asp	Phe	Ala	Ile	Asn	Glu	Ala	Val	Phe
	355						360					365			
Glu	Gln	Asp	Leu	Leu	Thr	Asp	Val	Gln	Cys	Ala	Phe	His	Val	Gln	Pro
	370					375					380				
Pro	Thr	Leu	Val	His	Ser	Tyr	Leu	Asn	Ile	His	Tyr	Val	Cys	Glu	Thr
385					390					395					400
Gly	Ser	Arg	Ile	Ile	Phe	Leu	Thr	Ile	His	Thr	Leu	Arg	Lys	Val	Pro
				405					410					415	
Val	Phe	Glu	Gln	Leu	Glu	Ala	His	Thr	Gln	Val	Lys	Leu	Leu	Arg	Gly
			420					425					430		
Val	Trp	Pro	Ala	Leu	Met	Ala	Ile	Ala	Leu	Ala	Gln	Cys	Gln	Gly	Gln
		435					440					445			
Leu	Ser	Val	Pro	Thr	Ile	Ile	Gly	Gln	Phe	Ile	Gln	Ser	Thr	Arg	Gln
	450					455					460				
Leu	Ala	Asp	Ile	Asp	Lys	Ile	Glu	Pro	Leu	Lys	Ile	Ser	Lys	Met	Ala
465					470					475					480
Asn	Leu	Thr	Arg	Thr	Leu	His	Asp	Phe	Val	Gln	Glu	Leu	Gln	Ser	Leu
				485					490					495	
Asp	Val	Thr	Asp	Met	Glu	Phe	Gly	Leu	Leu	Arg	Leu	Ile	Leu	Leu	Phe
			500				505						510		
Asn	Pro	Thr	Leu	Leu	Gln	Gln	Arg	Lys	Glu	Arg	Ser	Leu	Arg	Gly	Tyr
		515					520					525			
Val	Arg	Arg	Val	Gln	Leu	Tyr	Ala	Leu	Ser	Ser	Leu	Arg	Arg	Gln	Gly
	530					535					540				
Gly	Ile	Gly	Gly	Gly	Glu	Glu	Arg	Phe	Asn	Val	Leu	Val	Ala	Arg	Leu
545					550					555					560
Leu	Pro	Leu	Ser	Ser	Leu	Asp	Ala	Glu	Ala	Met	Glu	Glu	Leu	Phe	Phe
				565					570					575	
Ala	Asn	Leu	Val	Gly	Gln	Met	Gln	Met	Asp	Ala	Leu	Ile	Pro	Phe	Ile
			580					585					590		

Leu Met Thr Ser Asn Thr Ser Gly Leu  
595 600

<210> 26

<211> 2288

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence; note =  
synthetic construct

<400> 26

attggaacaa	ggagatttta	ttgcgttaga	aaaggttcaa	aataggcaca	aagtgcctga	60
aaatatcgta	actgaccgga	agtaacataa	ctttaaccaa	gtgcctcgaa	aaatagatgt	120
ttttaaaagc	tcaagaatgg	tgataacaga	cgtccaataa	gaattttcaa	agagccaaat	180
gtttgggttt	cagttattta	tacagccgac	gactatTTTT	tagccgcctg	ctgtggcgac	240
aatggacggc	gttaaggttg	agacgttcat	caaaagcgaa	gaaaaccgag	cgatgccctt	300
gatcggagga	ggcagtgcct	caggcggcac	tcctctgcc	ggaggcggcg	tgggaatggg	360
agccggagca	tccgcaacgt	tgagcgtgga	gctgtgtttg	gtgtgcgggg	accgcgcctc	420
cgggcggcac	tacggagcca	taagctgcga	aggctgcaag	ggattcttca	agcgctcgat	480
ccggaagcag	ctgggctacc	agtgtcgcgg	ggctatgaac	tgcgagggtca	ccaagcacca	540
caggaatcgg	tgccagttct	gtcgactaca	gaagtgcctg	gccagcggca	tgccaagtga	600
ttctgtgcag	cacgagagga	aaccgattgt	ggacaggaag	gaggggatca	tcgctgctgc	660
cggtagctca	tccacttctg	gcggcggtaa	tggctcgtcc	acctacctat	ccggcaagtc	720
cggctatcag	caggggctg	gcaaggggca	cagtgtaaag	gccgaatccg	cggccacgcc	780
tccagtgcac	agcgcgccag	caacggcctt	caatttgaat	gagaatatat	tcccgatggg	840
tttgaatttc	gcagaactaa	cgcagacatt	gatgttcgct	acccaacagc	agcagcaaca	900
acagcaacag	catcaacaga	gtggtagcta	ttcgccagat	attccgaagg	cagatcccga	960
ggatgacgag	gacgactcaa	tggacaacag	cagcacgctg	tgcttgcagt	tgctcgccaa	1020
cagcgccagc	aacaacaact	cgcagcacct	gaactttaat	gctgggggaag	taccacccgc	1080
tctgcctacc	acctcgacaa	tggggccttat	tcagagtctg	ctggacatgc	gggtcatcca	1140
caagggactg	cagatcctgc	agcccatcca	aaaccaactg	gagcgaaatg	gtaatctgag	1200
tgtgaagccc	gagtgcgatt	cagaggcgga	ggacagtggc	accgaggatg	ccgtagacgc	1260
ggagctggag	cacatggaac	tagactttga	gtgcggtggg	aaccgaagcg	gtggaagcga	1320
ttttgctatc	aatgaggcgg	tctttgaaca	ggatcttctc	accgatgtgc	agtgtgcctt	1380
tcattgtgcaa	ccgccgactt	tgggtccactc	gtattttaaat	attcattatg	tgtgtgagac	1440
gggctcgcga	atcatttttc	tcaccatcca	tacccttcga	aaggttccag	ttttcgaaca	1500
attggaagcc	catacacagg	tgaactcct	gagaggagtg	tggccagcat	taatggctat	1560
agctttggcg	cagtgtcagg	gtcagctttc	ggtgcccacc	attatcgggc	agtttattca	1620
aagcactcgc	cagctagcgg	atatcgataa	gatcgaaccg	ttgaagatct	cgaagatggc	1680
aaatctcacc	aggaccctgc	acgactttgt	ccaggagctc	cagtcactgg	atgttactga	1740
tatggagttt	ggcttgctgc	gtctgatctt	gctcttcaat	ccaacgctct	tgcaagcagc	1800
caaggagcgg	tcgttgcgag	gctacgtccg	cagagtccaa	ctctacgctc	tgtcaagttt	1860
gagaaggcag	ggtggcatcg	gcggcggcga	ggagcgcttt	aatgttcttg	tggctcgcct	1920
tcttccgctc	agcagcctgg	acgcagaggc	catggaggag	ctgttcttcg	ccaacttggt	1980
ggggcagatg	cagatggatg	ctcttatctc	gttcatactg	atgaccagca	acaccagtgg	2040
actgtaggcg	gaattgagaa	gaacagggcg	caagcagatt	cgctagactg	cccaaaagca	2100
agactgaaga	tggaccaagt	gcgggcaata	catgtagcaa	ctaggcaaata	cccatataatt	2160
atatatttaa	tatatacaat	atatagttta	ggatacaata	ttctaacata	aaaccatggg	2220
tttattgttg	ttcacagata	aaatggaatc	gatttcccaa	taaaagcgaa	tatgttttta	2280
aacagaat						2288

<210> 27

<211> 508

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence; note =  
synthetic construct

```

<400> 27
Met Asp Asn Cys Asp Gln Asp Ala Ser Phe Arg Leu Ser His Ile Lys
 1          5          10          15
Glu Glu Val Lys Pro Asp Ile Ser Gln Leu Asn Asp Ser Asn Asn Ser
 20          25          30
Ser Phe Ser Pro Lys Ala Glu Ser Pro Val Pro Phe Met Gln Ala Met
 35          40          45
Ser Met Val His Val Leu Pro Gly Ser Asn Ser Ala Ser Ser Asn Asn
 50          55          60
Asn Ser Ala Gly Asp Ala Gln Met Ala Gln Ala Pro Asn Ser Ala Gly
 65          70          75          80
Gly Ser Ala Ala Ala Ala Val Gln Gln Gln Tyr Pro Pro Asn His Pro
 85          90          95
Leu Ser Gly Ser Lys His Leu Cys Ser Ile Cys Gly Asp Arg Ala Ser
100          105          110
Gly Lys His Tyr Gly Val Tyr Ser Cys Glu Gly Cys Lys Gly Phe Phe
115          120          125
Lys Arg Thr Val Arg Lys Asp Leu Thr Tyr Ala Cys Arg Glu Asn Arg
130          135          140
Asn Cys Ile Ile Asp Lys Arg Gln Arg Asn Arg Cys Gln Tyr Cys Arg
145          150          155          160
Tyr Gln Lys Cys Leu Thr Cys Gly Met Lys Arg Glu Ala Val Gln Glu
165          170          175
Glu Arg Gln Arg Gly Ala Arg Asn Ala Ala Gly Arg Leu Ser Ala Ser
180          185          190
Gly Gly Gly Ser Ser Gly Pro Gly Ser Val Gly Gly Ser Ser Ser Gln
195          200          205
Gly Gly Gly Gly Gly Gly Gly Val Ser Gly Gly Met Gly Ser Gly Asn
210          215          220
Gly Ser Asp Asp Phe Met Thr Asn Ser Val Ser Arg Asp Phe Ser Ile
225          230          235          240
Glu Arg Ile Ile Glu Ala Glu Gln Arg Ala Glu Thr Gln Cys Gly Asp
245          250          255
Arg Ala Leu Thr Phe Leu Arg Val Gly Pro Tyr Ser Thr Val Gln Pro
260          265          270
Asp Tyr Lys Gly Ala Val Ser Ala Leu Cys Gln Val Val Asn Lys Gln
275          280          285
Leu Phe Gln Met Val Glu Tyr Ala Arg Met Met Pro His Phe Ala Gln
290          295          300
Val Pro Leu Asp Asp Gln Val Ile Leu Leu Lys Ala Ala Trp Ile Glu
305          310          315          320
Leu Leu Ile Ala Asn Val Ala Trp Cys Ser Ile Val Ser Leu Asp Asp
325          330          335
Gly Gly Ala Gly Gly Gly Gly Gly Gly Leu Gly His Asp Gly Ser Phe
340          345          350
Glu Arg Arg Ser Pro Gly Leu Gln Pro Gln Gln Leu Phe Leu Asn Gln
355          360          365
Ser Phe Ser Tyr His Arg Asn Ser Ala Ile Lys Ala Gly Val Ser Ala
370          375          380
Ile Phe Asp Arg Ile Leu Ser Glu Leu Ser Val Lys Met Lys Arg Leu
385          390          395          400
Asn Leu Asp Arg Arg Glu Leu Ser Cys Leu Lys Ala Ile Ile Leu Tyr
405          410          415
Asn Pro Asp Ile Arg Gly Ile Lys Ser Arg Ala Glu Ile Glu Met Cys
420          425          430
Arg Glu Lys Val Tyr Ala Cys Leu Asp Glu His Cys Arg Leu Glu His
435          440          445
Pro Gly Asp Asp Gly Arg Phe Ala Gln Leu Leu Leu Arg Leu Pro Ala
450          455          460

```

Leu Arg Ser Ile Ser Leu Lys Cys Gln Asp His Leu Phe Leu Phe Arg  
 465 470 475 480  
 Ile Thr Ser Asp Arg Pro Leu Glu Glu Leu Phe Leu Glu Gln Leu Glu  
 485 490 495  
 Ala Pro Pro Pro Pro Gly Leu Ala Met Lys Leu Glu  
 500 505

<210> 28

<211> 2488

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence; note =  
 synthetic construct

<400> 28

aaaaatgtcg	acgcgaaaaa	aggtatattat	tcattagtca	gaaagtctgg	cattctttgt	60
ttgttggttaa	aaagcgcaat	tgtttgagg	cgagcggaata	aagtgcgctg	ctccatcggc	120
tcaagattat	gtaaatgcag	caacgacccc	accaacaacg	aaactgcaac	ctgctccact	180
tggcccaacg	gaccaatagc	ggacggacgg	acacgggtggc	gttggcaaag	tgaaacccca	240
acagagaggc	gaaagcgagc	caagacacac	cacatacaca	cgaagagaac	gagcaagaag	300
aaaccggtag	gcggaggagg	cgctgcccc	agttcctcca	atatacccag	caccacatca	360
caagcccagg	atggacaact	gcgaccagga	cgccagcttt	cggtctgagcc	acatcaagga	420
ggaggtcaag	ccggacatct	cgcagctgaa	cgacagcaac	aacagcagct	tttcgcccaa	480
ggccgagagt	cccgtgccct	tcatgcaggc	catgtccatg	gtccacgtgc	tgcccggctc	540
caactccgcc	agctccaaca	acaacagcgc	tggagatgcc	caaattggcg	aggcgcccaa	600
ttcggctgga	ggctctgcgc	ccgctgcagt	ccagcagcag	tatccgccta	accatccgct	660
gagcggcagc	aagcacctct	gctctatattg	cggggatcgg	gccagtggca	agcactacgg	720
cgtgtacagc	tgtgagggct	gcaagggctt	ctttaaacgc	acagtgcgca	aggatctcac	780
atacgtttgc	agggagaacc	gcaactgcat	catagacaag	cggcagagga	accgctgcca	840
gtactgccgc	taccagaagt	gcctaacctg	cggcatgaag	cgcgaaagcg	tccaggagga	900
gcgtcaacgc	ggcgcccgc	atgcggcggg	taggctcagc	gccagcggag	gcggcagtag	960
cgggtccaggt	tcggtaggcg	gatccagctc	tcaaggcgga	ggaggaggag	gcggcgtttc	1020
tggcggaatg	ggcagcggca	acggttctga	tgacttcatg	accaatagcg	tgtccaggga	1080
tttctcgatc	gagcgcacat	tagaggccga	gcagcgagcg	gagacccaat	gcggcgatcg	1140
tgcactgacg	ttcctgcgcg	ttggtcccta	ttccacagtc	cagccggact	acaaggggtgc	1200
cgtgtcggcc	ctgtgccaa	tggtcaacaa	acagctcttc	cagatggctc	aatacgcgcg	1260
catgatgccg	cactttgccc	aggtgcccgt	ggacgaccag	gtgattctgc	tgaaagccgc	1320
ttggatcgag	ctgctcattg	cgaacgtggc	ctggtgcagc	atcgtttcgc	tggtatgacg	1380
cgggtgccggc	ggcggggggc	gtggactagg	ccacgatggc	tcctttgagc	gacgatcacc	1440
gggccttcag	cccagcagc	tgttcctcaa	ccagagcttc	tcgtaccatc	gcaacagtgc	1500
gatcaaagcc	ggtgtgtcag	ccatcttcga	ccgcatattg	tcggagctga	gtgtaaagat	1560
gaagcggctg	aatctcgacc	gacgcgagct	gtcctgcttg	aaggccatca	tactgtacaa	1620
cccggacata	cgcgggatca	agagccgggc	ggagatcgag	atgtgccgcg	agaaggtgta	1680
cgcttgccctg	gacgagcact	gcgcctgga	acatccgggc	gacgatggac	gctttgcgca	1740
actgctgctg	cgtctgccc	ctttgcgac	gatcagcctg	aagtgccagg	atcacctgtt	1800
cctcttcgcg	attaccagcg	acgggcccgt	ggaggagctc	tttctcgagc	agctggaggc	1860
gccgcgcgca	cccggcctgg	cgatgaaact	ggagtaggg	cccgactcta	aagtctcccc	1920
cgttctccat	ccgaaaaatg	tttcattgtg	attgcgtttg	tttgcatttc	tcctctctat	1980
cccttatata	ctacaaaagc	cccctaatat	tacgcaaaat	gtgtatgtaa	ttgtttatatt	2040
tttttttatt	acctaataat	attattatta	ttgatataga	aaatgttttc	cttaagatga	2100
agattagcct	cctcgacgtt	tatgtcccag	taaacgaaaa	acaaacaaaa	tccaaaactt	2160
gaaaagaaca	caaaacacga	acgagaaaa	gcacacaagc	aaagtaaaa	taaaagttaa	2220
actaaagcta	aacgagtaaa	gatattaaaa	taacgggttaa	aattaatgca	tagttatgat	2280
ctacagacgt	atgtaaacat	acaaattcag	cataaatata	tatgtcagca	ggcgcatatc	2340
tgcgggtgctg	gccccgttct	aaatcaattg	taattacttt	ttaacataaa	tttaccctaaa	2400
acgttatcaa	ttagatgcga	gatacaaaaa	tcaccgacga	aaaccaacaa	aatatatcta	2460
tgtataaaaa	atataaactg	cataacaa				2488

&lt;210&gt; 29

&lt;211&gt; 906

&lt;212&gt; PRT

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence; note =  
synthetic construct

&lt;400&gt; 29

Met	Gly	Glu	Glu	Leu	Pro	Ile	Leu	Lys	Gly	Ile	Leu	Lys	Gly	Asn	Val	1	5	10	15
Asn	Tyr	His	Asn	Ala	Pro	Val	Arg	Phe	Gly	Arg	Val	Pro	Lys	Arg	Glu	20	25	30	
Lys	Ala	Arg	Ile	Leu	Ala	Ala	Met	Gln	Gln	Ser	Thr	Gln	Asn	Arg	Gly	35	40	45	
Gln	Gln	Arg	Ala	Leu	Ala	Thr	Glu	Leu	Asp	Asp	Gln	Pro	Arg	Leu	Leu	50	55	60	
Ala	Ala	Val	Leu	Arg	Ala	His	Leu	Glu	Thr	Cys	Glu	Phe	Thr	Lys	Glu	65	70	75	80
Lys	Val	Ser	Ala	Met	Arg	Gln	Arg	Ala	Arg	Asp	Cys	Pro	Ser	Tyr	Ser	85	90	95	
Met	Pro	Thr	Leu	Leu	Ala	Cys	Pro	Leu	Asn	Pro	Ala	Pro	Glu	Leu	Gln	100	105	110	
Ser	Glu	Gln	Glu	Phe	Ser	Gln	Arg	Phe	Ala	His	Val	Ile	Arg	Gly	Val	115	120	125	
Ile	Asp	Phe	Ala	Gly	Met	Ile	Pro	Gly	Phe	Gln	Leu	Leu	Thr	Gln	Asp	130	135	140	
Asp	Lys	Phe	Thr	Leu	Leu	Lys	Ala	Gly	Leu	Phe	Asp	Ala	Leu	Phe	Val	145	150	155	160
Arg	Leu	Ile	Cys	Met	Phe	Asp	Ser	Ser	Ile	Asn	Ser	Ile	Ile	Cys	Leu	165	170	175	
Asn	Gly	Gln	Val	Met	Arg	Arg	Asp	Ala	Ile	Gln	Asn	Gly	Ala	Asn	Ala	180	185	190	
Arg	Phe	Leu	Val	Asp	Ser	Thr	Phe	Asn	Phe	Ala	Glu	Arg	Met	Asn	Ser	195	200	205	
Met	Asn	Leu	Thr	Asp	Ala	Glu	Ile	Gly	Leu	Phe	Cys	Ala	Ile	Val	Leu	210	215	220	
Ile	Thr	Pro	Asp	Arg	Pro	Gly	Leu	Arg	Asn	Leu	Glu	Leu	Ile	Glu	Lys	225	230	235	240
Met	Tyr	Ser	Arg	Leu	Lys	Gly	Cys	Leu	Gln	Tyr	Ile	Val	Ala	Gln	Asn	245	250	255	
Arg	Pro	Asp	Gln	Pro	Glu	Phe	Leu	Ala	Lys	Leu	Leu	Glu	Thr	Met	Pro	260	265	270	
Asp	Leu	Arg	Thr	Leu	Ser	Thr	Leu	His	Thr	Glu	Lys	Leu	Val	Val	Phe	275	280	285	
Arg	Thr	Glu	His	Lys	Glu	Leu	Leu	Arg	Gln	Gln	Met	Trp	Ser	Met	Glu	290	295	300	
Asp	Gly	Asn	Asn	Ser	Asp	Gly	Gln	Gln	Asn	Lys	Ser	Pro	Ser	Gly	Ser	305	310	315	320
Trp	Ala	Asp	Ala	Met	Asp	Val	Glu	Ala	Ala	Lys	Ser	Pro	Leu	Gly	Ser	325	330	335	
Val	Ser	Ser	Thr	Glu	Ser	Ala	Asp	Leu	Asp	Tyr	Gly	Ser	Pro	Ser	Ser	340	345	350	
Ser	Gln	Pro	Gln	Gly	Val	Ser	Leu	Pro	Ser	Pro	Pro	Gln	Gln	Gln	Pro	355	360	365	
Ser	Ala	Leu	Ala	Ser	Ser	Ala	Pro	Leu	Leu	Ala	Ala	Thr	Leu	Ser	Gly	370	375	380	
Gly	Cys	Pro	Leu	Arg	Asn	Arg	Ala	Asn	Ser	Gly	Ser	Ser	Gly	Asp	Ser	385	390	395	400

Gly	Ala	Ala	Glu	Met	Asp	Ile	Val	Gly	Ser	His	Ala	His	Leu	Thr	Gln
				405					410					415	
Asn	Gly	Leu	Thr	Ile	Thr	Pro	Ile	Val	Arg	His	Gln	Gln	Gln	Gln	Gln
			420					425						430	
Gln	Gln	Gln	Gln	Ile	Gly	Ile	Leu	Asn	Asn	Ala	His	Ser	Arg	Asn	Leu
			435				440						445		
Asn	Gly	Gly	His	Ala	Met	Cys	Gln	Gln	Gln	Gln	Gln	His	Pro	Gln	Leu
	450					455					460				
His	His	His	Leu	Thr	Ala	Gly	Ala	Ala	Arg	Tyr	Arg	Lys	Leu	Asp	Ser
465					470				475						480
Pro	Thr	Asp	Ser	Gly	Ile	Glu	Ser	Gly	Asn	Glu	Lys	Asn	Glu	Cys	Lys
				485				490						495	
Ala	Val	Ser	Ser	Gly	Gly	Ser	Ser	Ser	Cys	Ser	Ser	Pro	Arg	Ser	Ser
			500					505					510		
Val	Asp	Asp	Ala	Leu	Asp	Cys	Ser	Asp	Ala	Ala	Ala	Asn	His	Asn	Gln
		515					520					525			
Val	Val	Gln	His	Pro	Gln	Leu	Ser	Val	Val	Ser	Val	Ser	Pro	Val	Arg
	530					535					540				
Ser	Pro	Gln	Pro	Ser	Thr	Ser	Ser	His	Leu	Lys	Arg	Gln	Ile	Val	Glu
545					550				555						560
Asp	Met	Pro	Val	Leu	Lys	Arg	Val	Leu	Gln	Ala	Pro	Pro	Leu	Tyr	Asp
				565				570						575	
Thr	Asn	Ser	Leu	Met	Asp	Glu	Ala	Tyr	Lys	Pro	His	Lys	Lys	Phe	Arg
			580					585						590	
Ala	Leu	Arg	His	Arg	Glu	Phe	Glu	Thr	Ala	Glu	Ala	Asp	Ala	Ser	Ser
		595					600					605			
Ser	Thr	Ser	Gly	Ser	Asn	Ser	Leu	Ser	Ala	Gly	Ser	Pro	Arg	Gln	Ser
	610				615					620					
Pro	Val	Pro	Asn	Ser	Val	Ala	Thr	Pro	Pro	Pro	Ser	Ala	Ala	Ser	Ala
625					630					635					640
Ala	Ala	Gly	Asn	Pro	Ala	Gln	Ser	Gln	Leu	His	Met	His	Leu	Thr	Arg
			645					650						655	
Ser	Ser	Pro	Lys	Ala	Ser	Met	Ala	Ser	Ser	His	Ser	Val	Leu	Ala	Lys
			660					665					670		
Ser	Leu	Met	Ala	Glu	Pro	Arg	Met	Thr	Pro	Glu	Gln	Met	Lys	Arg	Ser
		675					680					685			
Asp	Ile	Ile	Gln	Asn	Tyr	Leu	Lys	Arg	Glu	Asn	Ser	Thr	Ala	Ala	Ser
	690				695						700				
Ser	Thr	Thr	Asn	Gly	Val	Gly	Asn	Arg	Ser	Pro	Ser	Ser	Ser	Ser	Thr
705				710						715					720
Pro	Pro	Pro	Ser	Ala	Val	Gln	Asn	Gln	Gln	Arg	Trp	Gly	Ser	Ser	Ser
				725				730						735	
Val	Ile	Thr	Thr	Thr	Cys	Gln	Gln	Arg	Gln	Gln	Ser	Val	Ser	Pro	His
			740					745					750		
Ser	Asn	Gly	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser
		755					760						765		
Ser	Ser	Ser	Ser	Thr	Ser	Ser	Asn	Cys	Ser	Ser	Ser	Ser	Ala	Ser	Ser
	770					775						780			
Cys	Gln	Tyr	Phe	Gln	Ser	Pro	His	Ser	Thr	Ser	Asn	Gly	Thr	Ser	Ala
785					790					795					800
Pro	Ala	Ser	Ser	Ser	Ser	Gly	Ser	Asn	Ser	Ala	Thr	Pro	Leu	Leu	Glu
				805				810						815	
Leu	Gln	Val	Asp	Ile	Ala	Asp	Ser	Ala	Gln	Pro	Leu	Asn	Leu	Ser	Lys
			820					825					830		
Lys	Ser	Pro	Thr	Pro	Pro	Pro	Ser	Lys	Leu	His	Ala	Leu	Val	Ala	Ala
		835					840					845			
Ala	Asn	Ala	Val	Gln	Arg	Tyr	Pro	Thr	Leu	Ser	Ala	Asp	Val	Thr	Val
	850					855					860				
Thr	Ala	Ser	Asn	Gly	Gly	Pro	Pro	Ser	Ala	Ala	Ala	Ser	Pro	Ala	Pro
865					870				875						880

<400> 30						
agtcacccgtc	gcagtcgcag	cagttgaggt	tgcctctcct	cgatttcggg	caaatccgat	60
accatatagc	acagcgtacc	gcactctggg	tatattcgta	acgcgccttg	gcttttacag	120
ttagtcgcgt	tcgagacctt	gtcgaagttt	gtcatgttag	ccagcgatcc	gcgggatccg	180
aaataagcca	agaatcacia	cgcgagtgcg	gcagttgcca	gcagtaacta	caccaatatt	240
tatattaatt	aaaataaatt	aatgaaaca	acatgctgat	taatgccaat	gaatgttaaa	300
tgcaattggt	aatgtgaaga	aaagtcgacc	aagtctcccc	aaaacaacac	ttattcaaca	360
tccactacac	actcgccttt	ctggattacg	cgcccaaaaa	aaaacaaaaa	ttaaaaatta	420
aaccaaacca	acaactaatt	tatttgctaa	atatccaaa	aattcaatca	atgtgaaaag	480
caagcaaaca	aagttcctct	cacaacaaaa	cagcagttaa	ttaaaatatc	taaccgagat	540
aaagtgcaa	gaagataaca	agtttctcaa	gcaaacatcc	atatgtacct	gagtaccaac	600
caaaaagctg	tgtgtgtgcc	aaaaaccgaa	gaggaattat	ccaaaaatat	ttaatgagca	660
agctcaactg	agtggttgat	gtgcccccca	agggaaaagt	gaccaagtca	agatattttg	720
tcaaatcgaa	cacagaaaac	acaaaaatgg	gcgaagaact	cccgatattg	aagggcatac	780
ttaaaggcaa	cgtcaactat	cacaatgcgc	ctgtgcgttt	tggacgcgtg	ccgaagcgcg	840
aaaaggcgcg	tatcctggcg	gccatgcaac	agagcaccca	gaatcgcggc	cagcagcgag	900
ccctcgcac	cgagctggat	gaccagccac	gcctcctcgc	cgccgtgctg	cgcgcccacc	960
tcgagacctg	tgagttcacc	aaggagaagg	tctcggcgat	gcggcagcgg	gcgcgggatt	1020
gcccctccta	ctccatgcc	acacttctgg	cctgtccgct	gaaccccgcc	cctgaactgc	1080
aatcggagca	ggagttctcg	cagcgtttcg	cccacgtaat	tcgcggcgtg	atcgactttg	1140
ccggcatgat	tcctggcttc	cagctgctca	cccaggacga	taagttcacg	ctcctgaagg	1200
cgggactctt	cgacgcctcg	tttgtgcgcc	tgatctgcat	gtttgactcg	tcgataaact	1260
caatcatctg	tctaaatggc	caggtgatgc	gacgggatgc	gatccagaac	ggagccaatg	1320
cccgcttcct	ggtggactcc	accttcaatt	tcgcggagcg	catgaactcg	atgaacctga	1380
cagatgccga	gataggcctg	ttctgcgcc	tcgttctgat	tacgccgat	cgccccgggt	1440
tgcgcaacct	ggagctgac	gagaagatgt	actcgcgact	caagggctgc	ctgcagtaca	1500
ttgtcgccca	gaataggccc	gatcagcccc	agttcctggc	caagttgctg	gagacgatgc	1560
ccgatctgcg	cacctgagc	acctgcaca	ccgagaaact	ggtagttttc	cgcaccgagc	1620
acaaggagct	gctgcgccag	cagatgtggt	ccatggagga	cggcaacaac	agcgatggcc	1680
agcagaacaa	gtcgcctctg	ggcagctggg	cggatgccat	ggacgtggag	gcggccaaga	1740
gtccgccttg	ctcggtatcg	agcactgagt	ccgccgacct	ggactacggc	agtccgagca	1800
gttcgcagcc	acagggcgtg	tctctgcct	cgccgcctca	gcaacagccc	tcggctctgg	1860
ccagctcggc	tcctctgctg	gcggccaccc	tctccggagg	atgtcccctg	cgcacccggg	1920
ccaattccgg	ctccagcggg	gactccggag	cagctgagat	ggatatcggt	ggctcgcacg	1980
cacatctcac	ccagaacggg	ctgacaatca	cgccgattgt	gcgacaccag	cagcagcaac	2040
aacagcagca	gcagatcgga	atactcaata	atgcgcattc	ccgcaacttg	aatggggggac	2100
acgcgatgtg	ccagcaacag	cagcagcacc	cacaactgca	ccaccacttg	acagccggag	2160
ctgcccgccta	cagaaagcta	gattcgccca	cggattcggg	cattgagtcg	ggcaacgaga	2220
agaacgagtg	caaggcggtg	agttcggggg	gaagttcctc	gtgctccagt	ccgcgttcca	2280
gtgtggatga	tgcgctggac	tgcagcgatg	ccgcgcgcaa	tcacaatcag	gtggtgcagc	2340
atccgcagct	gagtgtggtg	tccgtgtcac	cagttcgtct	gccccagccc	tccaccagca	2400
gccatctgaa	gcgacagatt	gtggaggata	tgcccgtgct	gaagcgcgtg	ctgcaggctc	2460
cccctctgta	cgataccaac	tcgctgatgg	acgaggccta	caagccgcac	aagaaattcc	2520
gggcctctcg	gcacgcgag	ttcgagaccg	ccgaggcgga	tgccagcagt	tccacttccg	2580
gctcgaacag	cctgagtgcc	ggcagtcgcg	gacagagtcc	agtcccgaac	agtgtggcca	2640
cgcccccgcc	atcggcgggc	agcgcgcgcg	caggtaatcc	cgcccagagc	cagctgcaca	2700
tgcacctgac	ccgcagcagc	cccaaggcct	cgatggccag	ctcgcactcg	gtgctggcca	2760

```

agtctctcat ggccgagccg cgcattgacgc ccgagcagat gaagcgcagc gatattatcc 2820
aaaactactt gaagcgcgag aacagcacag cagccagcag caccaccaat ggcggtgggca 2880
accgcagtc cagcagcagc tccacaccgc cgccatcggc ggtccagaat cagcagcggt 2940
ggggcagcag ctcggtgatc accaccacct gccagcagcg ccagcagtc gtgtcgccgc 3000
acagcaacgg ttccagctcc agttcgagct ctagctccag ctccagttcg tcatectcct 3060
ccacatcctc caactgcagc tccagctcgg ccagcagctg ccagtatttc cagtcgcccgc 3120
actccaccag caacggcacc agtgcaccgg cgagctccag ttcgggatcg aacagcgcca 3180
cgcccctgct ggaactgcag gtggacattg ctgactcggc gcagcctctc aatttgtcca 3240
agaaatcgcc cagcccgccg cccagcaagc tgcacgctct ggtggccgcg gccaatgccg 3300
ttcaaaggta tcccacattg tccgcccagc tcacagtgc agcctccaat ggcggtcctc 3360
cgtcggcgccg ggcgagtcgg gcgcccagca gcagtccgcc ggcgagtgtg ggctccccc 3420
atccgggcct gagcgccgcc gtgcacaagg taatgctgga ggcgtaagag cgggaggagg 3480
taggtggttt tacgcggaga agtgggagag acagagactg ggagtggcag ttcagcgaag 3540
caggaagcag gatcacttgg agcggcgagg gttgaattaa attattttac catttaattg 3600
agacgtgtac aaagtttgaa agcaaaacca acatgcattg aattttaaac taatatttaa 3660
agcaacaaca acaaaaacaa ctacaagtta ttaattttaa aaacaaacaa acaaacaaac 3720
aacaacaaac ccaagcttga atggtattac 3750

```

&lt;210&gt; 31

&lt;211&gt; 392

&lt;212&gt; PRT

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence; note =  
synthetic construct

&lt;400&gt; 31

```

Met His Pro Ser His Leu Gln Gln Gln Gln Gln Gln His Leu Leu Gln
1      5      10      15
Gln Gln Gln Gln Gln Gln His Gln Pro Gln Leu Gln Gln His His Gln
20      25      30
Leu Gln Gln Gln Pro His Val Ser Gly Val Arg Val Lys Thr Pro Ser
35      40      45
Thr Pro Gln Thr Pro Gln Met Cys Ser Ile Ala Ser Ser Pro Ser Glu
50      55      60
Leu Gly Gly Cys Asn Ser Ala Asn Asn Asn Asn Asn Asn Asn Asn
65      70      75      80
Ser Ser Ser Gly Asn Ala Ser Gly Gly Ser Gly Val Ser Val Gly Val
85      90      95
Val Val Val Gly Gly His Gln Gln Leu Val Gly Gly Ser Met Val Gly
100     105     110
Met Ala Gly Met Gly Thr Asp Ala His Gln Val Gly Met Cys His Asp
115     120     125
Gly Leu Ala Gly Thr Ala Asn Glu Leu Thr Val Tyr Asp Val Ile Met
130     135     140
Cys Val Ser Gln Ala His Arg Leu Asn Cys Ser Tyr Thr Glu Glu Leu
145     150     155     160
Thr Arg Glu Leu Met Arg Arg Pro Val Thr Val Pro Gln Asn Gly Ile
165     170     175
Ala Ser Thr Val Ala Glu Ser Leu Glu Phe Gln Lys Ile Trp Leu Trp
180     185     190
Gln Gln Phe Ser Ala Arg Val Thr Pro Gly Val Gln Arg Ile Val Glu
195     200     205
Phe Ala Lys Arg Val Pro Gly Phe Cys Asp Phe Thr Gln Asp Asp Gln
210     215     220
Leu Ile Leu Ile Lys Leu Gly Phe Phe Glu Val Trp Leu Thr His Val
225     230     235     240
Ala Arg Leu Ile Asn Glu Ala Thr Leu Thr Leu Asp Asp Gly Ala Tyr
245     250     255

```

Leu Thr Arg Gln Gln Leu Glu Ile Leu Tyr Asp Ser Asp Phe Val Asn  
 260 265 270  
 Ala Leu Leu Asn Phe Ala Asn Thr Leu Asn Ala Tyr Gly Leu Ser Asp  
 275 280 285  
 Thr Glu Ile Gly Leu Phe Ser Ala Met Val Leu Leu Ala Ser Asp Arg  
 290 295 300  
 Ala Gly Leu Ser Glu Pro Lys Val Ile Gly Arg Ala Arg Glu Leu Val  
 305 310 315 320  
 Ala Glu Ala Leu Arg Val Gln Ile Leu Arg Ser Arg Ala Gly Ser Pro  
 325 330 335  
 Gln Ala Leu Gln Leu Met Pro Ala Leu Glu Ala Lys Ile Pro Glu Leu  
 340 345 350  
 Arg Ser Leu Gly Ala Lys His Phe Ser His Leu Asp Trp Leu Arg Met  
 355 360 365  
 Asn Trp Thr Lys Leu Arg Leu Pro Pro Leu Phe Ala Glu Ile Phe Asp  
 370 375 380  
 Ile Pro Lys Ala Asp Asp Glu Leu  
 385 390

&lt;210&gt; 32

&lt;211&gt; 3341

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence; note =  
 synthetic construct

&lt;400&gt; 32

aagcattaac	gaaagaactg	cgcacaaagt	agggaggcaa	taattacata	tgtacatggc	60
tgggaaaggc	cttaactaaa	cttagcaaac	taataaatag	aaaaaaggaa	atattggcca	120
aatattatag	tattgggaat	attagggttac	ttgatataca	aaattaatgt	ctattttata	180
cacttattct	tagacttaat	gttaacttat	cgtacttatt	atgattgggt	tttcaagatt	240
accagaactt	gatagattgg	tctagctttt	gaaatcggat	agcattttct	ttaaaggact	300
ttgccatatg	ctaaagccta	acttcttttt	tcaattcagc	cacagctgac	aaaagcgaag	360
aaaatttgaa	agaccgtgaa	tccttttgaa	acgccctctc	cggattcctc	attaagtgca	420
aaagatataa	catcgcagag	atttcccata	aaaatgctga	tcaggcgccc	tcgcagggtg	480
ccaacgtcga	tttccgccag	caggacgatg	atgaagatga	tggatgccca	tctcaccgat	540
tcgatccgag	caacatggat	gtataccaaa	tagagctgga	ggaacaggca	caaataccgt	600
ccaaactgct	ggtcgaaacc	tgtgtgaagc	actcgtcttc	ggagcagcag	cagctccaag	660
ttaagcagga	ggacctcatc	aaggatttca	ctcgggacga	ggaggaacag	ccaagcgaag	720
aggaggcgga	ggaagaggac	aacgaagagg	acgaggaaga	agaaggcgaa	gaagaagagg	780
aggacgagga	cgaggaagcc	ctgctgccgg	tagtcaattt	taatgcaa	tcagacttta	840
atattgcattt	ctttgacaca	ccggaggact	cgtccaccca	aggggcctac	agtgaggcca	900
atagcttgga	atccgagcag	gaagaggaga	agcaaacaca	gcagcatcag	cagcagaagc	960
agcatcaccg	ggatttgag	gattgcctaa	gtgccattga	agctgatcca	ttgcagttgt	1020
tgcattgcga	cgacttctat	agaacatcag	ccctagcaga	gagtgttgca	gccagtctaa	1080
gccacagca	gcagcagcaa	cggcagcaca	cccaccagca	acaacagcaa	cagcagcagc	1140
agcagcaaca	ccctggacag	cagcaacatc	agctcaactg	cacgctgagc	aatggtggag	1200
gtgctttgta	caccatcagc	agtgtgcac	agttcgggtc	ggccagcaac	cacaacacca	1260
gcagcagctc	cccctcctcc	agcgccgccc	actcttcgcc	ggacagcggc	tgctcgtcgg	1320
cctcctcctc	cggatcttcg	cgatcctgcg	gatecctcct	tgcatecctc	tcctcgtcag	1380
cggtcagcag	caccatcagc	agcgcccgca	gcagcaacaa	cagcgtcgtc	aaccccgcag	1440
caacatcttc	atctgttgcg	catctgaaca	aagagcaaca	gcagcagcca	ctgccgacga	1500
cacagctgca	acagcagcag	cagcaccagc	agcagttgca	acacccgcag	cagcagcaat	1560
cttttggcct	agcagacagc	agcagcagca	acggcagcag	caacaacaac	aacggtgtct	1620
cctcgaaatc	atattgtgcc	tgcaaaagtct	gtggcgacaa	ggcatcgggg	taccactatg	1680
gtgtaacctc	ctgcgagggt	tgcaagggat	tctttcgtcg	cagtatccag	aagcaaatacg	1740
aatatcgctg	tttgcgggac	ggcaagtgcc	tggtcatcag	actgaaccgc	aatcgctgcc	1800
agtactgccg	cttcaagaaa	tgcccttccg	ctggcatgag	ccgcgattcc	gtacgttatg	1860
gtcgcgttcc	caagcggtcc	cgtgagctga	acggagcggc	cgcctcctcc	gccgccgctg	1920

```

gagctcctgc ctccctcaat gtggatgact ctaccagcag cacactgcac ccgagtcacc 1980
tacagcagca gcagcaacag catctactac agcagcaaca gcagcagcaa catcagccac 2040
agctgcagca acaccaccaa ctgcaacagc agccgcatgt aagcggcgta cgtgtgaaga 2100
ccccgagtac tccacaaacg ccacaaatgt gttcgatcgc ctctctcgcca tcggagctgg 2160
gcggttgcaa tagtgccaat aacaataaca ataataacaa caacagtagc agcggtaatg 2220
ccagcgggtgg cagcggcggtg agcgtcggcg ttgttggtgt gggcggacac cagcaactgg 2280
tgggaggcag catggtggga atggcgggca tgggcacgga tgcccaccag gtgggcatgt 2340
gtcacgacgg cttggcgggga acggcaaacg agctgaccgt ctacgatgtc atcatgtgcg 2400
tgtcgcaggc gcaccgcctc aactgctcct acacggagga actgaccaga gagctcatgc 2460
gtcgtcccgt gacggtgcca caaaatggga ttgccagcac agtggccgag agtctggagt 2520
tccagaagat ctggctgtgg caacagtctt cggccagggt gacgcctggc gttcagcgga 2580
ttgtggagtt tgcgaaacgc gtacctggct tctgtgattt caccgaagat gaccagctta 2640
tactaataaa gctgggcttc ttcgaggtct gggtgaccca tgtggcccgg ttgatcaatg 2700
aggcgacatt gacactggac gatggtgcct acctgacgcg ccagcagctt gagatactct 2760
acgattctga ctttgtcaac gccttgctga actttgcca cagctgaac gcctacgggc 2820
tgagtgcac cgaaatcgga ctcttctcgg ccatggtgct gcttgccctc gatcgagctg 2880
gactcagcga gcccaagggtg atcggcaggg ccagggaact ggtggccgag gcgctgcgcg 2940
tacagatcct gcgttcgcgg gcaggatccc cacaggcgct gcagctgatg ccggcgctgg 3000
aagccaagat acccgagctg agatccttgg gggccaagca cttctcacac ctagactggc 3060
tacggatgaa ctggaccaag ctgcgcctgc cgcccctctt cgccgagatc ttcgacatcc 3120
cgaaggctga cgatgagctg taggatgtgg agccaacccc gcgattccag ggccgtgcaa 3180
agcaaaccgc aacaagaaca gaatattcta ccacttgtag gcttaagcaa cgtagctata 3240
gatcgaaatg ggagggccgc agatcagata cacgtctact cagcattacc ggagagatag 3300
tccactaagc ctatatgcat actactatac tagcagtgtt a 3341

```

&lt;210&gt; 33

&lt;211&gt; 878

&lt;212&gt; PRT

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence; note =  
synthetic construct

&lt;400&gt; 33

```

Met Lys Arg Arg Trp Ser Asn Asn Gly Gly Phe Met Arg Leu Pro Glu
1          5          10          15
Glu Ser Ser Ser Glu Val Thr Ser Ser Ser Asn Gly Leu Val Leu Pro
20          25          30
Ser Gly Val Asn Met Ser Pro Ser Ser Leu Asp Ser His Asp Tyr Cys
35          40          45
Asp Gln Asp Leu Trp Leu Cys Gly Asn Glu Ser Gly Ser Phe Gly Gly
50          55          60
Ser Asn Gly His Gly Leu Ser Gln Gln Gln Gln Ser Val Ile Thr Leu
65          70          75          80
Ala Met His Gly Cys Ser Ser Thr Leu Pro Ala Gln Thr Thr Ile Ile
85          90          95
Pro Ile Asn Gly Asn Ala Asn Gly Asn Gly Gly Ser Thr Asn Gly Gln
100         105         110
Tyr Val Pro Gly Ala Thr Asn Leu Gly Ala Leu Ala Asn Gly Met Leu
115         120         125
Asn Gly Gly Phe Asn Gly Met Gln Gln Gln Ile Gln Asn Gly His Gly
130         135         140
Leu Ile Asn Ser Thr Thr Pro Ser Thr Pro Thr Thr Pro Leu His Leu
145         150         155         160
Gln Gln Asn Leu Gly Gly Ala Gly Gly Gly Gly Ile Gly Gly Met Gly
165         170         175
Ile Leu His His Ala Asn Gly Thr Pro Asn Gly Leu Ile Gly Val Val
180         185         190
Gly Gly Gly Gly Gly Val Gly Leu Gly Val Gly Gly Gly Val Gly
195         200         205

```

Gly	Leu	Gly	Met	Gln	His	Thr	Pro	Arg	Ser	Asp	Ser	Val	Asn	Ser	Ile
210						215					220				
Ser	Ser	Gly	Arg	Asp	Asp	Leu	Ser	Pro	Ser	Ser	Ser	Leu	Asn	Gly	Tyr
225					230					235					240
Ser	Ala	Asn	Glu	Ser	Cys	Asp	Ala	Lys	Lys	Ser	Lys	Lys	Gly	Pro	Ala
				245					250					255	
Pro	Arg	Val	Gln	Glu	Glu	Leu	Cys	Leu	Val	Cys	Gly	Asp	Arg	Ala	Ser
			260					265					270		
Gly	Tyr	His	Tyr	Asn	Ala	Leu	Thr	Cys	Glu	Gly	Cys	Lys	Gly	Phe	Phe
		275					280					285			
Arg	Arg	Ser	Val	Thr	Lys	Ser	Ala	Val	Tyr	Cys	Cys	Lys	Phe	Gly	Arg
	290					295					300				
Ala	Cys	Glu	Met	Asp	Met	Tyr	Met	Arg	Arg	Lys	Cys	Gln	Glu	Cys	Arg
305					310					315					320
Leu	Lys	Lys	Cys	Leu	Ala	Val	Gly	Met	Arg	Pro	Glu	Cys	Val	Val	Pro
				325					330					335	
Glu	Asn	Gln	Cys	Ala	Met	Lys	Arg	Arg	Glu	Lys	Lys	Ala	Gln	Lys	Glu
			340					345					350		
Lys	Asp	Lys	Met	Thr	Thr	Ser	Pro	Ser	Ser	Gln	His	Gly	Gly	Asn	Gly
	355						360					365			
Ser	Leu	Ala	Ser	Gly	Gly	Gly	Gln	Asp	Phe	Val	Lys	Lys	Glu	Ile	Leu
	370					375					380				
Asp	Leu	Met	Thr	Cys	Glu	Pro	Pro	Gln	His	Ala	Thr	Ile	Pro	Leu	Leu
385					390					395					400
Pro	Asp	Glu	Ile	Leu	Ala	Lys	Cys	Gln	Ala	Arg	Asn	Ile	Pro	Ser	Leu
				405					410					415	
Thr	Tyr	Asn	Gln	Leu	Ala	Val	Ile	Tyr	Lys	Leu	Ile	Trp	Tyr	Gln	Asp
		420						425					430		
Gly	Tyr	Glu	Gln	Pro	Ser	Glu	Glu	Asp	Leu	Arg	Arg	Ile	Met	Ser	Gln
	435						440					445			
Pro	Asp	Glu	Asn	Glu	Ser	Gln	Thr	Asp	Val	Ser	Phe	Arg	His	Ile	Thr
	450					455				460					
Glu	Ile	Thr	Ile	Leu	Thr	Val	Gln	Leu	Ile	Val	Glu	Phe	Ala	Lys	Gly
465					470					475					480
Leu	Pro	Ala	Phe	Thr	Lys	Ile	Pro	Gln	Glu	Asp	Gln	Ile	Thr	Leu	Leu
				485					490					495	
Lys	Ala	Cys	Ser	Ser	Glu	Val	Met	Met	Leu	Arg	Met	Ala	Arg	Arg	Tyr
			500					505					510		
Asp	His	Ser	Ser	Asp	Ser	Ile	Phe	Phe	Ala	Asn	Asn	Arg	Ser	Tyr	Thr
	515						520					525			
Arg	Asp	Ser	Tyr	Lys	Met	Ala	Gly	Met	Ala	Asp	Asn	Ile	Glu	Asp	Leu
	530					535					540				
Leu	His	Phe	Cys	Arg	Gln	Met	Phe	Ser	Met	Lys	Val	Asp	Asn	Val	Glu
545					550					555					560
Tyr	Ala	Leu	Leu	Thr	Ala	Ile	Val	Ile	Phe	Ser	Asp	Arg	Pro	Gly	Leu
				565					570					575	
Glu	Lys	Ala	Gln	Leu	Val	Glu	Ala	Ile	Gln	Ser	Tyr	Tyr	Ile	Asp	Thr
			580					585					590		
Leu	Arg	Ile	Tyr	Ile	Leu	Asn	Arg	His	Cys	Gly	Asp	Ser	Met	Ser	Leu
	595					600						605			
Val	Phe	Tyr	Ala	Lys	Leu	Leu	Ser	Ile	Leu	Thr	Glu	Leu	Arg	Thr	Leu
	610					615					620				
Gly	Asn	Gln	Asn	Ala	Glu	Met	Cys	Phe	Ser	Leu	Lys	Leu	Lys	Asn	Arg
625					630					635					640
Lys	Leu	Pro	Lys	Phe	Leu	Glu	Glu	Ile	Trp	Asp	Val	His	Ala	Ile	Pro
				645					650					655	
Pro	Ser	Val	Gln	Ser	His	Leu	Gln	Ile	Thr	Gln	Glu	Glu	Asn	Glu	Arg
			660					665					670		
Leu	Glu	Arg	Ala	Glu	Arg	Met	Arg	Ala	Ser	Val	Gly	Gly	Ala	Ile	Thr
	675						680					685			

Ala Gly Ile Asp Cys Asp Ser Ala Ser Thr Ser Ala Ala Ala Ala Ala  
690 695 700  
Ala Gln His Gln Pro Gln Pro Gln Pro Gln Pro Gln Pro Ser Ser Leu  
705 710 715 720  
Thr Gln Asn Asp Ser Gln His Gln Thr Gln Pro Gln Leu Gln Pro Gln  
725 730 735  
Leu Pro Pro Gln Leu Gln Gly Gln Leu Gln Pro Gln Leu Gln Pro Gln  
740 745 750  
Leu Gln Thr Gln Leu Gln Pro Gln Ile Gln Pro Gln Pro Gln Leu Leu  
755 760 765  
Pro Val Ser Ala Pro Val Pro Ala Ser Val Thr Ala Pro Gly Ser Leu  
770 775 780  
Ser Ala Val Ser Thr Ser Ser Glu Tyr Met Gly Gly Ser Ala Ala Ile  
785 790 795 800  
Gly Pro Ile Thr Pro Ala Thr Thr Ser Ser Ile Thr Ala Ala Val Thr  
805 810 815  
Ala Ser Ser Thr Thr Ser Ala Val Pro Met Gly Asn Gly Val Gly Val  
820 825 830  
Gly Val Gly Val Gly Gly Asn Val Ser Met Tyr Ala Asn Ala Gln Thr  
835 840 845  
Ala Met Ala Leu Met Gly Val Ala Leu His Ser His Gln Glu Gln Leu  
850 855 860  
Ile Gly Gly Val Ala Val Lys Ser Glu His Ser Thr Thr Ala  
865 870 875

<210> 34

<211> 5586

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence; note =  
synthetic construct

<400> 34

tagtattttt	ttggactttg	ttgttaacgg	ttgttcgctc	gcacgtacga	agcccgatcg	60
cgttcgtcaa	aaaacaagat	acaaaataca	gcacacacaa	ttgaaaacga	caacctaaaca	120
gtacgggtttc	ccaaagcacc	ttacattttca	aaaccgaaaa	cccccaaat	gttgtaacca	180
aataatgttt	aatcacata	tacacctaca	tatatattatg	aaaaattgtt	agacaaatcc	240
caaataatac	cagttcccc	aacaaccgca	acaaacacaa	gtgcaattca	tcggcaaaaa	300
ttaatataaa	gtgcaaatgc	attgtagctg	aaactcaaac	aatagtaaaa	atacatatat	360
aagtgggtgaa	gaagcaaaag	gaaatagttc	ttaaaataac	gcaaatcgag	agcatatat	420
catattttgta	cagatattat	atggcggtctg	catagtgcga	actgcggctg	aggggaataca	480
gcggtatcga	aatgtaaata	ggaaacaacg	aagccagaac	tcgaaatcaa	acatcagcaa	540
cgtgacacac	agacataaga	cgcccgtcta	gtcgtggtct	gtggaacgct	agctccgctt	600
tgccaggagc	cggagacttt	ttccgcatcc	acaatattac	atatgtacat	atatcgaaga	660
tagtgcgcgga	gtgagtggag	gattttgtgcc	gtggatcccg	atcccccttac	atatatataa	720
aggtagtga	aagattttac	tcaacattcc	aaatagtgtc	ttgtcaactg	gaataccttt	780
tggtcaaata	cgcagtgggc	ccatggatac	ttgtggatta	gtagcagaac	tggcgacta	840
tatcgacgca	tatgctctga	ttgtttcccg	cactaaatga	gcagggattc	gggcgaaaat	900
gtatttttgaa	cgcaaacaag	tgcgcaaaaa	atactagctc	caccacgaaa	ctgcacaaaa	960
caccgccaga	agcgagcaga	acctcgggcc	gcacgaccga	gcttcgtaaa	gcaacagagg	1020
atcttaccag	gagatagctc	ttctccacat	agaccaactg	ccagggacaa	gctccttgct	1080
cccagccgac	gctaagtga	cggaaaacgg	ccacaaaacg	gcgactatcg	gctgccagag	1140
gatgaagcgg	cgctgggtcga	acaacggcgg	cttcatgcgc	ctaccggagg	agtcgtcctc	1200
ggaggtcacg	tcctcctcga	acgggctcgt	cctgccctcg	gggggtgaaca	tgtcgccctc	1260
gtcgtcggac	tcgcacgact	attgcgatca	ggacctttgg	ctctgcggca	acgagtcagg	1320
ttcgttttggc	ggctccaacg	gccatggcct	aagtcagcag	cagcagagcg	tcatcacgct	1380
ggccatgcac	gggtgctcca	gcactctgcc	cgcgcagaca	accatcatte	cgatcaacgg	1440
caacgcgaat	gggaatggag	gctccacca	tggccaatat	gtgccgggtg	ccactaatct	1500
gggagcgttg	gccaacggga	tgctcaatgg	gggcttcaat	ggaatgcagc	aacagattca	1560

gaatggccac	ggcctcatca	actccacaac	gccctcaacg	ccgaccaccc	cgctccacct	1620
tcagcagaac	ctggggggcg	cgggcgggcg	cggtatcggg	ggaatgggta	ttcttcacca	1680
cgcgaatggc	accccaaagt	gccttatcgg	agttgtggga	ggcggcgggc	gagtaggtct	1740
tggagtaggc	ggaggcggag	tgggaggcct	gggaatgcag	cacacacccc	gaagcgattc	1800
ggtgaattct	atatcttcag	gtcgcgatga	tctctcgcct	tcgagcagct	tgaacggata	1860
ctcggcgaac	gaaagctgcg	atgcgaagaa	gagcaagaag	ggacctgcgc	cacgggtgca	1920
agaggagctg	tgcctgggtt	gcggcgacag	ggcctccggc	taccactaca	acgccctcac	1980
ctgtgagggc	tgcaaggggt	tctttcgcag	cagcgttacg	aagagcgccg	tctactgctg	2040
caagttcggg	cgcgcctgcg	aatggacat	gtacatgagg	cgaaagtgtc	aggagtgcgc	2100
cctgaaaaag	tgcctggccg	tgggtatgcg	gccggaatgc	gtcgtcccgg	agaaccaatg	2160
tgcgatgaag	cggcgcgaaa	agaaggccca	gaaggagaag	gacaaaatga	ccacttcgcc	2220
gagctctcag	catggcggca	atggcagctt	ggcctctggt	ggcggccaag	actttgttaa	2280
gaaggagatt	cttgacctta	tgacatgcga	gccgccccag	catgccacta	ttccgctact	2340
acctgatgaa	atattggcca	agtgtcaagc	gcgcaatata	ccttccttaa	cgtacaatca	2400
gttggcgctt	atatacaagt	taattttggta	ccaggatggc	tatgagcagc	catctgaaga	2460
ggatctcagg	cgtataatga	gtcaaccgca	tgagaacgag	agccaaacgg	acgtcagctt	2520
tcggcatata	accgagataa	ccatactcac	ggtccagttg	attgttgagt	ttgctaaagg	2580
tctaccagcg	tttacaaga	taccccagga	ggaccagatc	acgttactaa	aggcctgctc	2640
gtcggaggtg	atgatgctgc	gtatggcacg	acgctatgac	cacagctcgg	actcaatatt	2700
cttcgcgaat	aatagatcat	atacgcgga	ttcttacaaa	atggccggaa	tggctgataa	2760
cattgaagac	ctgctgcatt	tctgccgcca	aatgttctcg	atgaagggtg	acaacgtcga	2820
atacgcgctt	ctcactgcc	ttgtgatctt	ctcggaccgg	ccgggcctgg	agaaggccca	2880
actagtgcga	gcatccaga	gctactacat	cgacacgcta	cgcatttata	tactcaaccg	2940
ccactgcggc	gactcaatga	gcctcgtctt	ctacgcaaag	ctgctctcga	tcctcaccga	3000
gctgcgtacg	ctgggcaacc	agaacgccga	gatgtgtttc	tcactaaagc	tcaaaaaccg	3060
caaactgccc	aagttcctcg	aggagatctg	ggacgttcat	gccatcccgc	catcgggtcca	3120
gtcgcacctt	cagattaccc	aggaggagaa	cgagcgtctc	gagcgggctg	agcgtatgcg	3180
ggcatcgggt	gggggcgcca	ttaccgccgg	cattgattgc	gactctgcct	ccacttcggc	3240
ggcggcagcc	gcggcccagc	atcagcctca	gcctcagccc	cagccccaac	cctcctccct	3300
gacccagaac	gattcccagc	accagacaca	gccgcagcta	caacctcagc	taccacctca	3360
gctgcaaggt	caactgcaac	cccagctcca	accacagctt	cagacgcaac	tccagccaca	3420
gattcaacca	cagccacagc	tccttcccgt	ctccgctccc	gtgcccgcct	ccgtaaccgc	3480
acctggttcc	ttgtccgcgg	tcagtacgag	cagcgaatac	atgggcggaa	gtgcggccat	3540
aggacccatc	acgcgcggca	ccaccagcag	tatcacggct	gccgttaccc	ctagctccac	3600
cacatcagcg	gtaccgatgg	gcaacggagt	tggagtccgt	gttgggggtg	gcggcaacgt	3660
cagcatgtat	gcgaacgccc	agacggcgat	ggccttgatg	ggtgtagccc	tgcattcgca	3720
ccaagagcag	cttatcgggg	gagtggcggt	taagtccggag	cactcgacga	ctgcatagca	3780
ggcgcagagt	cagctccacc	aacatcacca	ccacaacatc	gacgtcctgc	tggagtagaa	3840
agcgcagctg	aaccacacaca	gacatagggg	aatgggggaa	gttctctcca	gagagttcga	3900
gccgaactaa	atagtaaaaa	gtgaataatt	aatggacaag	cgtaaaaatgc	agttatttag	3960
tcttaagcct	gcaaataatta	cctattattc	atacaaatga	acataataata	cagcctatta	4020
acaattacgc	taaagcttaa	ttgaaaaagc	ttcaacaaca	attggacaaa	cgcgttgagg	4080
aaccgggaga	aaattttaaga	aaaaaaaaaac	cattgaaaat	tatgaaattt	agtatacatt	4140
ttttttgggt	ggatgtatgt	cgcacacagc	tcacgatcaa	ttctcgaatt	ttgttaacta	4200
aattgatcct	ccaaactgca	tgcgaaacag	atcagaaaag	agaacagaca	gtagggcggtg	4260
aacagaggga	agagagaaga	gaataaagat	tgttttatatt	taaaaaatat	ataaaataat	4320
aattactaac	tctaaacgta	atgaaagcaa	ctgtataata	tctaactata	actataaatt	4380
cgtactgtag	ggaagtgaga	aatctgttta	aatgaaacaa	aaataatgat	aataacatta	4440
tcacccacca	taattaaaat	cattttaaagt	aattaaaaac	aaaacacttt	taaaacacgc	4500
aaaacttgga	ctgatttttat	aaatatTTTT	taatcataaa	gaaaggcaac	ctgaaaaaaa	4560
tattacaaaa	acaaataaca	acatatTTTT	ttatgacacc	cttatatggt	ttcaaaacga	4620
gaattttaaat	tcttagattc	ttataatttt	atccaaaaat	attagccagc	aaaaaccttt	4680
attattggca	ttgttttttag	acatgttttt	aaaaaaaaact	ttgatattga	aactaaacaa	4740
aggataatga	aatgaaagtg	attggagtct	tactcaaaaa	ccaaaaggca	tcaaaaggta	4800
ttaaatttaa	aatataatct	aatttcgagt	tcaagaaaca	cttttttggtg	gaaaatagtt	4860
ttcaatcact	ttgataaaaa	ccacacaaat	taataaatac	atgcatacac	caaaagactt	4920
caatatatat	tttttaaaatt	tacattgata	attcgaaatt	tgaataagaa	tcacatccat	4980
ctaattttggc	taaatcaaaa	tttttatgaa	agccacacaa	aaaacgtgca	aattttgatta	5040
ctttggcaat	ttttatgtta	tacaaaattt	atgcaattga	ttttcaaaat	aattttttatt	5100
agattgtatt	agtttcattt	tgctttggga	tgtacatttt	aaataaattt	tacttttaaat	5160
tgttggcctt	atttttaactt	aaatcaaaat	tatttctaatt	ttagtaaaaa	aaaatgtggt	5220

```

taaaattgaa aataagaaca ctgtaaaata ttaataaaaa attaaagttt aaagtgattc 5280
ttttattatg taaaaagaag acaaaaaata tcttacgtag ctttctactt gaattgtgca 5340
atTTTTtact ttactacta atcctaattt aaatataatt tacacacacg cctacacatc 5400
cagccacata tttttaattt taagtcaacc taatttataa atatgaattt gtataatgac 5460
gaactaaaat tagcatgaca tcatggacat acttggaaat aactctatca aacgagctaa 5520
atgcattgaa gaagaaaatt cttgttaa atagtctgca cttcgacaaa cgaaaatcag 5580
tgaatt 5586

```

&lt;210&gt; 35

&lt;211&gt; 808

&lt;212&gt; PRT

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence; note =  
synthetic construct

&lt;400&gt; 35

```

Met Pro Asn Met Ser Ser Ile Lys Ala Glu Gln Gln Ser Gly Pro Leu
 1          5          10          15
Gly Gly Ser Ser Gly Tyr Gln Val Pro Val Asn Met Cys Thr Thr Thr
          20          25          30
Val Ala Asn Thr Thr Thr Thr Leu Gly Ser Ser Ala Gly Gly Ala Thr
          35          40          45
Gly Ser Arg His Asn Val Ser Val Thr Asn Ile Lys Cys Glu Leu Asp
 50          55          60
Glu Leu Pro Ser Pro Asn Gly Asn Met Val Pro Val Ile Ala Asn Tyr
 65          70          75          80
Val His Gly Ser Leu Arg Ile Pro Leu Ser Gly His Ser Asn His Arg
          85          90          95
Glu Ser Asp Ser Glu Glu Glu Leu Ala Ser Ile Glu Asn Leu Lys Val
          100          105          110
Arg Arg Arg Thr Ala Ala Asp Lys Asn Gly Pro Arg Pro Met Ser Trp
          115          120          125
Glu Gly Glu Leu Ser Asp Thr Glu Val Asn Gly Gly Glu Glu Leu Met
          130          135          140
Glu Met Glu Pro Thr Ile Lys Ser Glu Val Val Pro Ala Val Ala Pro
 145          150          155          160
Pro Gln Pro Val Cys Ala Leu Gln Pro Ile Lys Thr Glu Leu Glu Asn
          165          170          175
Ile Ala Gly Glu Met Gln Ile Gln Glu Lys Cys Tyr Pro Gln Ser Asn
          180          185          190
Thr Gln His His Ala Ala Thr Lys Leu Lys Val Ala Pro Thr Gln Ser
          195          200          205
Asp Pro Ile Asn Leu Lys Phe Glu Pro Pro Leu Gly Asp Asn Ser Pro
          210          215          220
Leu Leu Ala Ala Arg Ser Lys Ser Ser Ser Gly Gly His Leu Pro Leu
 225          230          235          240
Pro Thr Asn Pro Ser Pro Asp Ser Ala Ile His Ser Val Tyr Thr His
          245          250          255
Ser Ser Pro Ser Gln Ser Pro Leu Thr Ser Arg His Ala Pro Tyr Thr
          260          265          270
Pro Ser Leu Ser Arg Asn Asn Ser Asp Ala Ser His Ser Ser Cys Tyr
          275          280          285
Ser Tyr Ser Ser Glu Phe Ser Pro Thr His Ser Pro Ile Gln Ala Arg
          290          295          300
His Ala Pro Pro Ala Gly Thr Leu Tyr Gly Asn His His Gly Ile Tyr
 305          310          315          320

```

Arg	Gln	Met	Lys	Val	Glu	Ala	Ser	Ser	Thr	Val	Pro	Ser	Ser	Gly	Gln	
				325					330					335		
Glu	Ala	Gln	Asn	Leu	Ser	Met	Asp	Ser	Ala	Ser	Ser	Asn	Leu	Asp	Thr	
			340					345					350			
Val	Gly	Leu	Gly	Ser	Ser	His	Pro	Ala	Ser	Pro	Ala	Gly	Ile	Ser	Arg	
		355					360					365				
Gln	Gln	Leu	Ile	Asn	Ser	Pro	Cys	Pro	Ile	Cys	Gly	Asp	Lys	Ile	Ser	
	370					375					380					
Gly	Phe	His	Tyr	Gly	Ile	Phe	Ser	Cys	Glu	Ser	Cys	Lys	Gly	Phe	Phe	
385					390					395					400	
Lys	Arg	Thr	Val	Gln	Asn	Arg	Lys	Asn	Tyr	Val	Cys	Val	Arg	Gly	Gly	
				405					410					415		
Pro	Cys	Gln	Val	Ser	Ile	Ser	Thr	Arg	Lys	Lys	Cys	Pro	Ala	Cys	Arg	
			420					425					430			
Phe	Glu	Lys	Cys	Leu	Gln	Lys	Gly	Met	Lys	Leu	Glu	Ala	Ile	Arg	Glu	
		435					440					445				
Asp	Arg	Thr	Arg	Gly	Gly	Arg	Ser	Thr	Tyr	Gln	Cys	Ser	Tyr	Thr	Leu	
	450					455					460					
Pro	Asn	Ser	Met	Leu	Ser	Pro	Leu	Leu	Ser	Pro	Asp	Gln	Ala	Ala	Ala	
465					470					475					480	
Ala	Ala	Ala	Ala	Ala	Ala	Val	Ala	Ser	Gln	Gln	Gln	Pro	His	Gln	Arg	
				485					490					495		
Leu	His	Gln	Leu	Asn	Gly	Phe	Gly	Gly	Val	Pro	Ile	Pro	Cys	Ser	Thr	
		500						505					510			
Ser	Leu	Pro	Ala	Ser	Pro	Ser	Leu	Ala	Gly	Thr	Ser	Val	Lys	Ser	Glu	
		515					520					525				
Glu	Met	Ala	Glu	Thr	Gly	Lys	Gln	Ser	Leu	Arg	Thr	Gly	Ser	Val	Pro	
	530					535					540					
Pro	Leu	Leu	Gln	Glu	Ile	Met	Asp	Val	Glu	His	Leu	Trp	Gln	Tyr	Thr	
545					550					555					560	
Asp	Ala	Glu	Leu	Ala	Arg	Ile	Asn	Gln	Pro	Leu	Ser	Ala	Phe	Ala	Ser	
				565					570					575		
Gly	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Gly	Thr	Ser	Ser	Gly	Ala	His	
			580					585					590			
Ala	Gln	Leu	Thr	Asn	Pro	Leu	Leu	Ala	Ser	Ala	Gly	Leu	Ser	Ser	Asn	
		595					600					605				
Gly	Glu	Asn	Ala	Asn	Pro	Asp	Leu	Ile	Ala	His	Leu	Cys	Asn	Val	Ala	
	610					615					620					
Asp	His	Arg	Leu	Tyr	Lys	Ile	Val	Lys	Trp	Cys	Lys	Ser	Leu	Pro	Leu	
625					630					635					640	
Phe	Lys	Asn	Ile	Ser	Ile	Asp	Asp	Gln	Ile	Cys	Leu	Leu	Ile	Asn	Ser	
				645					650					655		
Trp	Cys	Glu	Leu	Leu	Leu	Phe	Ser	Cys	Cys	Phe	Arg	Ser	Ile	Asp	Thr	
		660						665					670			
Pro	Gly	Glu	Ile	Lys	Met	Ser	Gln	Gly	Arg	Lys	Ile	Thr	Leu	Ser	Gln	
		675					680					685				
Ala	Lys	Ser	Asn	Gly	Leu	Gln	Thr	Cys	Ile	Glu	Arg	Met	Leu	Asn	Leu	
		690				695					700					
Thr	Asp	His	Leu	Arg	Arg	Leu	Arg	Val	Asp	Arg	Tyr	Glu	Tyr	Val	Ala	
705					710					715					720	
Met	Lys	Val	Ile	Val	Leu	Leu	Gln	Ser	Asp	Thr	Thr	Glu	Leu	Gln	Glu	
				725					730					735		
Ala	Val	Lys	Val	Arg	Glu	Cys	Gln	Glu	Lys	Ala	Leu	Gln	Ser	Leu	Gln	
		740						745					750			
Ala	Tyr	Thr	Leu	Ala	His	Tyr	Pro	Asp	Thr	Pro	Ser	Lys	Phe	Gly	Glu	
		755					760					765				
Leu	Leu	Leu	Arg	Ile	Pro	Asp	Leu	Gln	Arg	Thr	Cys	Gln	Leu	Gly	Lys	
	770					775					780					
Glu	Met	Leu	Thr	Ile	Lys	Thr	Arg	Asp	Gly	Ala	Asp	Phe	Asn	Leu	Leu	
785					790					795					800	

Met Glu Leu Leu Arg Gly Glu His  
805

<210> 36

<211> 4841

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence; note =  
synthetic construct

<400> 36

actaaca	aaaa	caaacattttt	gctacttctgt	cgcaggcggg	actgtgttgc	gtcgtgtgat	60
cgctagagcg	ggtgtggaat	cggattcgag	cgcaaaacac	cgttcacgct	gtgagcgaaa		120
aagagtggta	gcgcctacag	tggcatatgt	agttaaatacc	gtgaataagt	gaaaaatccg		180
atatttgtcg	tgcaataatt	tcctcgattg	gcatcaagtg	gcttccagtc	gggtacatat		240
tgcacaagaa	atgtttatacg	cataatgtgc	acgcaaatta	aacgaattct	ctatgaaaat		300
gtgactagaa	tgtgagtcga	acaaaacgag	taaaacgtga	aatcccaact	ggcttttggg		360
taacaaatct	tatcaacaca	gcaacggaaa	tacattaaaa	tcttgataga	ctgagaaagg		420
gacaattgga	atacttttag	ttatttttta	atgtttttaca	acacaatgga	actgcatcaa		480
cgacacctct	caaactttta	caaattgcac	aactgagaaa	tagtctttga	taaataaata		540
aaatataaga	aatcgctact	gaaacaagat	gccaaacatg	tccagcatca	aagcggagca		600
gcaaagcggg	cctcttggag	gaagtagcgg	ctatcaagta	ccgggtcaaca	tgtgcaccac		660
cacagtgcg	aatacgacga	ccactttggg	aagctccgcc	gggggagcca	ctggctccc		720
gcacaacgtc	tccgtgacaa	acatcaagtg	cgaactagac	gaactaccgt	caccgaacgg		780
caacatggtg	ccggttatcg	caaactacgt	tcacggtagc	ttgcgcattc	cactcagtgg		840
acattcaa	at	catagggagt	ccgattcggg	ggaggagctg	gcaagtattg	agaacttgaa	900
ggttcggcga	aggacggcgg	cggacaaaaa	tggtcctcgt	ccaatgtcct	gggagggcga		960
gctgagcgat	actgaggtca	acggggggcga	agagctgatg	gaaatggagc	caacaattaa		1020
gagtgaggtg	gtccctgctg	ttgcaccccc	acaaccgcgc	tgcgcactac	aaccgataaa		1080
aacagagcta	gagaacattg	caggcgagat	gcagattcaa	gagaagtgtt	acccccagtc		1140
caacacacaa	catcacgctg	ccacaaaatt	aaaagtggcc	ccgacgcaaa	gtgatccgat		1200
caatctcaag	ttcgaaccgc	ctctggggaga	caattctccg	ctactggctg	cacgtagcaa		1260
gtccagcagt	ggaggccacc	taccactgcc	aacgaatccc	agtcccgcact	ccgccataca		1320
ttccgtctac	acgcacagct	ccccctcgca	gtcgcctctg	acgtcgcgcc	acgcccccta		1380
cactccgtct	ctgagccgca	acaacagcga	cgcctcgcac	agtagctgct	acagctatag		1440
ctccgaattc	agtcccacac	actcgcacct	tcaagcgcgt	catgccccac	ccgccggcac		1500
gctctatggc	aaccacccatg	gtattttaccg	ccagatgaag	gtggaagcct	catccactgt		1560
gccgtccagt	gggcaggagg	cgcagaacct	gagtatggac	tctgcctcta	gcaatctgga		1620
tacagtgggc	ttaggatctt	cgcacccccgc	atctccggcg	ggcatatcac	gtcagcagtt		1680
gatcaactcg	ccctgccccca	tctgcgggtga	caagatcagc	ggattttcatt	acgggattttt		1740
ctcctgcgag	tcttgcaagg	gcttcttcaa	gcgcaccgtg	caaaatcgca	agaactacgt		1800
gtgcgtgcgt	ggtggaccat	gtcagggtcag	cattttccacg	cgcaagaaat	gtccagcctg		1860
ccgcttcgag	aagtgtctgc	agaagggaat	gaaactagaa	gcgattcggg	aggaccgaac		1920
ccgtggcggc	cgctccacat	accagtgtct	ctacacgctg	cccaactcaa	tgcttagtcc		1980
gctgcttagt	cctgatcaag	cggcagcagc	tgccgcgcga	gcagcagtgg	caagtcagca		2040
gcagccgcac	cagcgactac	atcaactaaa	tggattttgga	ggtgtaccca	ttccctgctc		2100
tactttctct	ccagccagcc	ctagtttggc	aggaacttcg	gtcaagtccg	aagagatggc		2160
ggagacgggc	aagcaaagcc	tccgaacggg	aagcgtacca	ccactactgc	aggaaatcat		2220
ggatgtagag	catctgtggc	agtacaccga	tgcagagctg	gcccgcacat	accaaccact		2280
gtccgcattc	gcctctggca	gctcttcgct	gtcgtcatcg	tcaggtagat	cctcaggcgc		2340
ccatgcacaa	ctcaccaatc	cactactggc	tagtgctggg	ctctcgtcca	atggcgagaa		2400
tgccaatcct	gatcttatcg	ctcatctctg	caacgtggct	gatcacccgc	tttataaaat		2460
cgtcaaatgg	tgcaagagct	tgccgcctttt	taagaacatt	tcgatcgatg	accaaattctg		2520
cttgctcatt	aactcgtggg	gcgagctgtt	gctcttctcc	tgctgtttta	gatcaattga		2580
tactcctgga	gagattaaaa	tgtcacaagg	caggaagata	accctatcgc	aggccaaatc		2640
aatggccttg	cagacttgca	ttgaacggat	gctcaacctc	acagatcacc	tgaggcgatt		2700
gcgcgttgat	cgctacgaat	atgttgccat	gaaagtattt	gtgctgttgc	agtcagatac		2760
gacagagtta	caggaagcgg	taaagggtgcg	cgagtgtcag	gaaaaagctt	tgcagagctt		2820
gcaagcttac	accctggcgc	attatcctga	cacgccatcc	aagttttggg	agctttttgct		2880

acgcattcct	gatttgcagc	gaacgtgcca	gcttggcaag	gagatgttga	cgatcaagac	2940
tcgcgatgga	gctgatttca	atttgctaata	ggagcttttg	cgcgagagac	attgacaatt	3000
gataactaag	acggaaatct	tttaccattg	gcaaaacaag	tttcacatat	ttagtattag	3060
atatatatat	tctatagata	agatccttac	tgtaagtctt	gaaaacatgt	gcctaaaaac	3120
caaagccacg	atagcagtca	catcaggccc	actggctcag	attaaatcca	agagcaagat	3180
tgccaaatth	ttacaccaat	atatatthtg	atatgagcca	tgtgcagggc	ctcagatcgc	3240
tgttgttgct	ggctaaagtt	tcagtaagaa	aagtatatat	tgattttgct	atttatacat	3300
atttgactta	tgtatagtgt	aaactaaagc	acacatggaa	aatgaaaaga	ctaaacaaat	3360
ttattttaaag	attactthta	ctattataga	aaaaggggaa	aaataaaaaa	cacaaaggca	3420
gagaagaaaa	tttagttaca	acaggtagcg	acattthttat	attthcttat	ataaggaaat	3480
attcaatgta	thtttaaata	aaagccaaac	ccgattthgg	ttgggaaaga	gctactgaaa	3540
thtttgatat	ctatatattc	atcactagaa	gacgaatgaa	tgtatccaat	gtthaaatgt	3600
tgtagcgtth	agthtttagtg	caattthcaca	catgtctaca	tacatgaata	ttcagcgaga	3660
tatgtthtgca	aactattata	aagcaaaaaga	ccactcgaaa	tcgccatcac	tggttggtgct	3720
aagactattc	cagttatgct	gtthgttgca	taaaaaacca	caactacgta	catcaataaa	3780
atgtataaatt	thtttatggga	gtthttagatt	tgtattaact	thttccttat	aattacgatt	3840
attattatta	ttactaathth	tatgaatatt	gtgtaacact	gacttaataa	gctgaaaaaa	3900
tcctgcaaca	ggattthaaaa	cacctgaata	cacaaaacat	tataacatga	atacatthtg	3960
cttatggcct	agatagthtg	atatgtactt	tgcatatgta	tgcatgtgtc	tatatgtgag	4020
tacgtaccat	acaaattcct	gtcccaccag	aaaaatcaca	cgcaataaaa	aattccaaaa	4080
tactaagctc	gtatctacaa	agaaagatta	aaagacaaat	tgatgaatag	gaatatgttg	4140
ccggaagtcc	aagagatttg	gctgaaagta	tcgacaaatt	ttcaacacat	cgttcatgga	4200
tattgtgcta	acactctcag	tttgaaaatc	attthctgtt	aaactthcta	tataataagt	4260
thtccattcg	attthgtatt	tacaattthgt	thctthtaatt	thctthtata	agttgtatct	4320
atgaaacatg	aggatctcag	thcatattga	tcgtgtthct	ctgccgtaca	ccgctthctgt	4380
ccgttaaatgt	aaaccataag	tataaatgaa	attagthtaa	tgthttattta	taaataaagc	4440
gctataataa	attthcaatac	attthataata	gttaactgat	taagaccact	gaaatcaaaa	4500
atattthtatt	tactaagcaa	agcacacgca	aacaatttht	aatgtthtatt	acgttaacaa	4560
caaactcatt	thtaataaatt	ctthtatgaat	acacaaagtt	acgcaattth	ccctctaggc	4620
gcattgctta	aatagthtaa	gaaaaataat	aaaccatag	cgcaatattth	aatgtaaaac	4680
agththctct	gcgtgtgatg	thtgctctag	ctacgtacaa	attcatcatt	ththaaattth	4740
aaaactcaat	thtgctthth	aataaattth	ataagtaaaa	thcaacaata	attgatatac	4800
aattgtcaat	gcaatattth	gtaataaaaa	tgcgaaaaat	c		4841

&lt;210&gt; 37

&lt;211&gt; 7555

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence; note =  
synthetic construct

&lt;400&gt; 37

gggccccccc	tcgaggctga	cggtatcgat	aagcttgccg	gtggcgagaga	aggtgtatcc	60
gtgattaaga	aagagccagc	cgatgagaag	cagccacagc	cacatgacca	cggtgcgtcc	120
gtacacaaga	tggcactgct	ggtgccgtth	cgagaccgat	ttgaggaaact	cctccagttc	180
gtccccca	tgaccgcctt	thtgaagcgg	cagggcggtg	cgaccacat	ctthgtgctg	240
aaccagggtg	acaggthccg	cttcaatcgc	gcctctctca	tcaacgtggg	thtccagtht	300
gccagcgatg	tgtacgatta	cattgccatg	cacgacgtag	acttgctgcc	cttgaatgac	360
aatctgctct	atgagtatcc	cagcagcttg	ggaccactgc	acatcgccgg	accgaagcta	420
catcccaa	accactatga	taacttcgtt	ggagggaata	tactggtgcg	acgcgagcac	480
thtaagcaga	tgaacggcat	gtcgaaccag	tactggggct	ggggattaga	ggacgacgag	540
thcttcgtgc	gcattccggga	tgccaggactg	caggtgacgc	ggccgcagaa	cattaagact	600
ggcactaatg	atacatthcag	gtgagaccag	tgctccggat	thcgcaacta	gacgtgacta	660
ctaataatta	thgtcattca	acctcagcca	tattcacaac	cgctatcata	gtaagcggga	720
caccagaag	tgctthcaacc	agaaggagat	gacccgcaag	cgggaccaca	agacgggcct	780
ggacaacgtg	aagtacaaaa	tacttaaggt	gcattgagatg	ctcattgacc	aggtgccggg	840
gaccatctc	aacattthtg	tcgattgtga	tgthtaataa	acgccttggt	gcgactgctc	900
cggaacggca	gcggctgcat	cgccggtaca	aacctgatgg	gttggtgttaa	accaaagatc	960
ctatgttht	thcgctatta	tagtgtgttg	tattgtataa	atgcgcta	acacgtgcac	1020

catgccatag	aggaatgtcc	agaagagcac	gtaggtgcaa	aggccgcccc	tgaactgatt	1080
ggtcagcaga	tttctgcggt	taatgaaaaa	cttgcgccac	tgggtgcccc	atttcacgag	1140
caccagaatc	cagagcacga	acacggacag	gaagtagaaa	aggaatccca	gcgtaccact	1200
caggcccaaa	atacctgcga	attgggtggga	cattaactaa	gttgggttcac	catcaattgg	1260
agccaattac	ccgcagcgca	gcccagagatg	gcagccatcg	atgtgcgaca	gtattccacg	1320
gcggatatgt	tgttccggat	ggcgccctcg	ctgtaggcga	ttatttcgcc	agtcttggac	1380
tgcgtgggtct	tcactcgatt	catttttattt	aattaaattc	tactttaatt	tctagcaaaa	1440
atattcctag	gctgtgaact	tcgatttgtgt	gccgatttgtg	ttatcgattg	gtgccgataa	1500
ctatgcactg	taaaaattca	ctagcgggttt	ttgcaggata	aatagttttt	gtaaattttc	1560
cgagataaac	ttgacgagct	gtttaatgtt	aaataatgaa	gtttaataca	atatcaaata	1620
tatttgctga	agtgtatatt	tattctcacc	gctctgtgct	tcgatggctc	acaattgcgt	1680
ttgccattcg	cccgggcacg	tagattgttg	ttattgggat	tggcctggag	cactcggacg	1740
gacagtaatt	cattaaaata	tgtgggtgata	acgcgagctg	ccgaatctgc	gtgcaattcg	1800
tgcgtttgac	gtgggtacta	actgctatgc	tgtcgcgcgg	acagttgttc	tgatacgag	1860
agttcctgcc	tcaccacaca	cgaccacctc	cattaaaacc	agccaccccc	cccagcgcct	1920
cctccaccga	cagcagctgc	tccaccgcac	caccaggaga	ggggcaatta	aaaaatcaat	1980
cagagggccc	atcacttgct	tgtaacccgc	gaagaactgc	gcggtgtgcg	gggacaaggc	2040
tctgggctac	aacttcaatg	cggtcacctg	cgagagctgc	aaggcgttct	tccgacggaa	2100
cgcgctggcc	aagaagcagt	tcacctgccc	cttcaaccaa	aactgcgaca	tcactgtggt	2160
cactcgacgc	ttctgccaga	aatgccgcct	gcgcaagtgc	ctggatatcg	ggatgaagag	2220
tgaaaacatt	atgtccgagg	aggacaagct	gatcaagcgg	cgcaagatcg	agaccaaccg	2280
ggccaagcga	cgcctcatgg	agaacggcac	ggatgcgtgc	gacgccgatg	gcggcgagga	2340
aagggatcac	aaagcgccgg	cggatagcag	cagcagcaac	cttgaccact	actcggggtc	2400
acaggactcg	cagagctgcg	gctcggcgga	cagcggggcc	aatgggtgct	ccggcagaca	2460
ggccagttcg	ccgggcacac	aggtcaatcc	gcttcagatg	acggccgaga	agatagtcga	2520
ccagatcgta	tccgaccggg	atcgagcctc	gcaggccatc	aaccggttga	tgcgcacgca	2580
gaaagaggct	atatcggtga	tggagaaggt	aatcagctca	caaaaggacg	ccttaaggct	2640
ggtgtcgcat	ttgatcgact	atccagggtg	gtgcagacaa	gatttcatcg	tttagcctta	2700
tccgctcacc	tatgaacgac	ttgaatcttt	acaggcgacg	cactcaagat	catttcaaag	2760
tttatgaact	cgccctttaa	cgcgctgaca	ggttagagtt	ttaaaatttg	tggtttttaa	2820
cttaattttca	cattccttgt	taattttaaat	acgcagtatt	caccaaattc	atgagctcac	2880
ccacggacgg	cgttgaaatt	atctdaaaga	tagttgattc	gcccgcggac	gtgggtggagt	2940
tcatgcagaa	cttgatgcac	tcgccagagg	acgccatcga	tataatgaac	aagttcatga	3000
ataccccagc	ggaggcgctg	cgcattctta	accgaatcct	aagcggcgga	ggagcgaacg	3060
cagcccagca	gacagcagac	cgcaagccat	tgctggacaa	ggagccggcg	gtgaagcctg	3120
cagcgccagc	ggagcgagct	gatactgtca	ttcaaagcat	gctgggcaac	agtccgcca	3180
tttcgccaca	tgatgctgcc	gtggatctgc	agtaccactc	gcccgggtgtc	ggggagcagc	3240
ccagtacatc	gagtagccac	cccttgccct	acatagccaa	ctcgccggac	ttcgatctga	3300
agaccttcat	gcagaccaac	tacaacgacg	agcccagttc	ggacagtgat	tttagcatta	3360
actcaatcga	atcggtgcta	tccgagggtga	tccgcattga	gtaccaggcc	ttcaatagca	3420
tacaacaagc	ggcatcgcg	gtaaaggagg	agatgtccta	cggcactcag	tctacgtacg	3480
gtggatgcaa	ttcggctgca	aacaatagcc	agccgcacct	gcagcaaccc	atctgcgccc	3540
catccaccca	gcagttggat	cgcgagctaa	acgaggcgga	gcaaatgaag	ctgcgggagc	3600
tgcgactggc	cagcgaggct	ctttatgatc	ccgtggacga	ggacctcagc	gccctgatga	3660
tgggcgatga	tcgcatttaag	gtaacccgct	agggataaca	gggtaataac	agtccacggt	3720
attagcctat	aggtctttct	acattttatag	ctccaacacc	acggcttata	taatcagagt	3780
gtgcgagctg	cgatatatgt	acacacggca	cctggcactt	tttagccatt	cgggtgattca	3840
gtgcgtctct	cgatgttggc	ccacggggccg	tatcttcgtc	agccagtttc	tgggttccca	3900
gcaatgctcg	cctaccaa	gtaaacacac	tttttaattg	gggtggctcaa	agttttttgat	3960
ttcccaagag	ctttggctga	gtaaaagaaa	attgatcgaa	ccagataagc	tatttttccc	4020
cagaggggtta	aagaatttga	agtcatgcga	ctgggtctag	ttaagatatt	tgattacgaa	4080
aattggcctt	taattaagac	cctaaacgtg	acaaacttcc	attctatata	cttcttgatg	4140
agtattttaa	caaataatggc	tatttttcgga	acaaatcggg	cactcattta	tatcttttagc	4200
tttatcttta	ttttttaaga	tgtgtccaca	cctttgatcg	acctctagtt	cccctggaga	4260
aatgatttgg	aattatccaa	taatgattca	tcacttccac	gaattgttgt	cccattaatc	4320
gagccaccct	agctttcatg	caatcagaac	gtctgggtctg	ccaagaagga	gcagacagcg	4380
gctttatcag	cctctggggcg	tgccaattgt	gacactatca	accattatca	agagtaccag	4440
caggcgctca	tgagtctcca	gccagccgtc	gattttgggcg	tttattgtcg	cttagactgt	4500
ttaccgattt	tgctcatcg	caattagcac	atttcagtat	tgtttaattgg	gaaaaacgat	4560
acaattttga	cgaaatatat	ggagcagcca	ggtgttgggc	gctatgataa	gcagtgtctc	4620
gccattcogat	tgagtcacct	tccagggaga	agcctttacg	attatggcga	taataatggc	4680

caccaaagag	aacatgggca	acatacgcac	tgacctgctc	aagtttgccg	aaggcaatat	4740
ctacgaggag	caccaaaggt	tcatacacaac	gtttgacgag	aagtggcgca	tggacgagaa	4800
cataatcctg	atcatgtgtg	ccattgtcct	ttttacctcg	gctcgatcgc	gagtgatata	4860
caaagacgtg	attagattgg	aacaggtgag	taagcacttg	ataccatact	gcagtattac	4920
taactttctt	tcattcgata	gaattcctac	tattatcttc	tgcaagata	tctggagagt	4980
gtttattctg	gctgtgaggc	gagaaacgcg	tttatcaagc	taatccaaaa	gatttcagat	5040
gtggagcgtc	tgaacaagtt	cataattaat	gtctatttga	atgttaaccc	atcccagggtg	5100
gagcccttgc	tgctgaaat	attcgatttg	aaaaatcact	agacaaccga	tgctgtcgg	5160
gcatttaattg	cctatgttga	tgcccaatga	tgaatgggtca	acaagctgta	gttgttgttg	5220
ttgttgatgt	ctgttttata	ttgtcgcttg	taatgttaga	ttttaatcga	atgtgattgt	5280
tagatttgca	tatactgcat	agattttata	tttctacatc	aaagagagca	tatttaggat	5340
accaagtgc	aagcaacaca	atctatatgt	aatgtacacc	gtttacctag	tttcaaataa	5400
actagacgat	aatgcaataa	ctaacttgga	agcgtgggtt	ctgtgcaaaa	aggaaaaaag	5460
acaaaaaaa	taaactgact	ttgagaacca	gtggtaataa	aatgtctcgt	attcttttct	5520
actcgaatga	atttcgaacc	ctccaggaca	aattacgcaa	acgagtgatt	ttgaacaaca	5580
atccaaaata	atttaattcc	gaaagtcaca	aaataaaaaat	tcgaagtagg	aaaaaaca	5640
taagatgttt	ggaaaccaac	gagagatgtg	cttcgttaaa	gcatcaaccc	ggggaaacac	5700
cacagcaacc	gcgcattgtg	acccgcgacc	agtcctcaga	aatccacgtc	gtgtacgtat	5760
ccgcagccag	cgtatgtgtc	cgcatctgcc	gaccccgctc	tacatagtca	tttatgtata	5820
atgtaggtaa	tataatagct	cgagctcgct	ccgcaccacc	aatgtgcgtc	gtgcaagtcc	5880
attccaattg	ttatccggtc	cactcgccgc	gcaaatacggc	tttcagggtg	attcgcgcca	5940
atccttgggc	cattgcagaa	actcatccaa	cgcgctgacg	gccaaattgc	gagaaagagc	6000
cttcacacgc	agattacgat	cggttgtaat	gagcaccaat	tcggtttgaa	tgaaacactt	6060
gccatctgca	aaagagtfff	agttagaaat	gctatcagga	aggacattta	acggaagcag	6120
ctcacctgtg	cattgttcgg	tttttgccgt	tttcgagaca	gccattgccg	ttgccagaat	6180
cttgtcgtca	ttggacaaat	attcctcctc	aactagggca	aacaccgatg	cattgacaaa	6240
cgagcccttt	gtggtagcac	atctagaaaa	gaaatcaata	aggtattatt	gatcagcagg	6300
aaaagctttc	ctgaacaact	attactactg	atttaaaaagt	aaaatttcaa	tacattatca	6360
ggaaactttt	atctatctca	atagcaacca	atgaattaga	cagaattata	aatagctaata	6420
cgctagtaaa	cccttttatca	gatatacagta	ataaaggaac	tatgagctga	cgcgcggaat	6480
ataatttaaca	atagcttact	tcacattgcc	tttggccgac	ttgatgaact	ctaacgactt	6540
tttggcccg	gacgacacct	cgtcaaagtg	gtggatgcgc	tgctctgct	tcgaactgcg	6600
gtacgagtcc	aacttaacgc	ccttgagag	gccatccagt	tccttaacca	ctgtcagtgg	6660
tataataagt	gtgtagcgtt	taaactccgt	ggacagtttt	tcaaagtctt	caaggcagtc	6720
gataaagcag	ttggtatccg	gtagaagata	gcgcggctgc	acctcgatgt	atattttcgt	6780
gtccacgaac	ttgagaatgt	cctccaactt	gctggtgcat	atctgcttaa	ccctagcctt	6840
ggcctcaagc	tctttcttca	gcttgacag	ctcggaaca	tcggaatcag	ttgaacgaca	6900
caacaattcc	tcaacggctt	tacattgaat	ctttgaaacg	ttcgccgcaa	gaccacaact	6960
ttcagcatca	tagttttcca	atgcgttctt	catcgctgcg	tttaggtgct	cgacagtga	7020
gcttgattcc	aacggctgct	ggagagcctc	ttgatgctgg	acgtagtatt	cctggaactg	7080
accaatccta	cgaacacgtt	cgaaaaactg	aagacgttcg	gatcctcttt	tgagatagtc	7140
catgctcagc	gtttcacgac	ccaacggagt	gaaacctcta	agggccacat	cctcatccaa	7200
cagtatttctg	tttttctcaa	gtttgtgctt	ccccattaag	cattcaatgt	actcaaaaag	7260
gattgttagc	tcagcccagc	aatcgatgaa	agagtgtctg	caatattgga	gttaatgaaa	7320
aactaataaa	aggcattcaa	tttatacata	ctcttccgat	ctgacgggtt	cccacacgtc	7380
caaactgatg	ctcaaccaac	gtacataaac	attcacgaac	tgtaggtatg	tgttcacagt	7440
cgcaaagctg	aaataaaaaga	ttaattagca	ataataaata	aacaaggcga	atttttagctt	7500
actcttcttc	tgggagacac	tggacatttg	tagaatcctc	tagatctact	agtcc	7555

&lt;210&gt; 38

&lt;211&gt; 545

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence; note =  
synthetic construct

&lt;400&gt; 38

gaagcaagcc	tctagaaaga	tgaagctact	gtcttctatc	gaacaagcat	gcgatatattg	60
ccgacttaaa	aagctcaagt	tcgcgatggc	ggcgaggaaa	gggatcacaa	agcgccggcg	120
gatagcagca	gcagcaacct	tgaccactac	tcggcagaaa	gaggctatat	cggtgatgga	180
gaaggtaatc	agctcacaaa	aggacgcctt	aacagaggac	gccatcgata	taatgaacaa	240
gttcatgaat	accccagctc	gcccgggtgc	ggggagcagc	ccagtacatt	ctacgtacgg	300
tggatgcaat	ctgaagttca	tcacaacggt	tgacgagaag	tggcgcatgg	acgagaacat	360
aatcctgata	atgtgtgcca	ttgtccttta	atgtctatatt	gaatgttaac	ccatcccagg	420
tggagccctt	gctgcgtgaa	atattcgatc	aaagagagca	tatttaggat	accaagtgca	480
aagcaacaca	atctataaga	cgataatgca	ataactaact	tggaagcgtg	ggttctgtgc	540
aaacc						545

&lt;210&gt; 39

&lt;211&gt; 1119

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence; note =  
synthetic construct

&lt;400&gt; 39

tggcgaatgg	gacgcgccct	gtagcggcgc	attaagcgcg	gcgggtgtgg	tggttacgcg	60
cagcgtgacc	gctacacttg	ttaggggtgat	ggttcttaaat	acaacctatt	aatttcccct	120
cgtcaaaaat	aaggttatca	agtgagaaat	caccatgagt	gacgactaac	cggcgcagga	180
acactgccag	cgcataca	atattttcac	ctgaatcagg	atatgcttcc	catacaatcg	240
atagattgtc	gcacctgatt	gcccgcagca	tcttcttgag	atcctttttt	tctgcgcgtt	300
ggcgataagt	cgtgtcttgg	tagtgagcga	ggaagcggaa	gagcgcctga	tgcggtatatt	360
tctccttacg	catctgtgcg	gtatttcaca	ccgcagggag	ctgcatgtgt	cagagggtttt	420
caccgtcatc	accgaaacgc	gcgaggcagc	tgccggcgatg	aaacgagaga	ggatgctcac	480
gatacggggt	actgatgatg	aaacgggaaac	cgaagaccat	tcatgtttgtt	gctcagaaga	540
ttccgaatac	cgcaagcgc	cactgtcttc	ggatcgtcgc	tatcccacta	ccgagatatc	600
cgcaccaacg	cgcagcccgg	actcggtaat	ggcgcgcatt	gccgagacag	aacttaattgg	660
gcccgcctaac	agcgcgattt	gctgggtgacc	caatgcgacc	agatcgcttt	acaggcttcg	720
acgcgccttc	gttctaccat	cgacaccacc	acgcttcacc	acgcgggaaa	cggtctgata	780
agagacaccg	gaaggagatg	gcgcccacaa	gtccctctag	aaataaaacc	ttgaccacta	840
ctcgggggtca	caggactcgc	agagctgcgg	ctcggcggac	agcggggcca	atgggtgctc	900
cggcacctta	aggctgggtg	cgcattttgat	cgactatcca	ggcgacgcac	tcaagatcat	960
ttcaaagttt	agctgcgcac	tcttaaccga	atcctaagcg	gcggaggagc	gaacgcagcc	1020
cagctacata	gccaactcgc	cggacttcga	tctgaagacc	ttcaagcaac	ccatctgcgc	1080
cccatccacc	cagcattccg	tgacaaacta	tatccggat			1119

&lt;210&gt; 40

&lt;211&gt; 30

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence; note =  
synthetic construct

&lt;400&gt; 40

gagagatgtg	cttcgttaaa	gcatcaaccc	30
------------	------------	------------	----

&lt;210&gt; 41

&lt;211&gt; 44

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence; note =  
synthetic construct

<400> 41  
ggactagtag atctagagga ttctacaaat gtccagtgtc tccc 44

<210> 42  
<211> 27  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence; note =  
synthetic construct

<400> 42  
ccattattat cgccataatc gtaaagg 27

<210> 43  
<211> 46  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence; note =  
synthetic construct

<400> 43  
attaccctgt tatccctagc gggttacctt aatgcgatca tcgccc 46

<210> 44  
<211> 30  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence; note =  
synthetic construct

<400> 44  
ggaaagcttt tcctgctgat caataatacc 30

<210> 45  
<211> 41  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence; note =  
synthetic construct

<400> 45  
tgggcccacac acttgcttgt aaccgccgaa gaactgcgcg g 41

<210> 46  
<211> 47  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence; note =  
synthetic construct

<400> 46  
cgctagggat aacagggtaa taacagtcca cggatttagc ctatagg 47

<210> 47  
<211> 47  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence; note =  
synthetic construct

<400> 47  
cgattatggc gataataatg gccaaagaga acatgggcaa catacgc 47

<210> 48  
<211> 26  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence; note =  
synthetic construct

<400> 48  
gaagcaagcc tctagaaaga tgaagc 26

<210> 49  
<211> 39  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence; note =  
synthetic construct

<400> 49  
cgtgccgttc tccatcgata cagtcaactg tctttgacc 39

<210> 50  
<211> 23  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence; note =  
synthetic construct

<400> 50  
gcctggatag tcgatcaa at gcg 23

<210> 51  
<211> 20  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence; note =  
synthetic construct

<400> 51  
atggagaacg gcacggatgc 20

<210> 52  
<211> 40  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence; note =  
synthetic construct

<400> 52  
tacattctag agaccaacta caacgacgag cccagtctgg 40

<210> 53  
<211> 41  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence; note =  
synthetic construct

<400> 53  
cattcatccg gacattaatt atgaacttgt tcagacgctc c 41

<210> 54  
<211> 39  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence; note =  
synthetic construct

<400> 54  
gggcatcaac tccggaatta aatgcccgcac acgcatcgg 39

<210> 55  
<211> 42  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence; note =  
synthetic construct

<400> 55  
gtctcacgac gttttgaacc cagaaatcga gctcgcccgg gg 42

<210> 56  
<211> 36  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence; note =  
synthetic construct

<400> 56

cacgaattcc aaactgtctc acgacgtttt gaaccc

36

<210> 57

<211> 44

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence; note =  
synthetic construct

<400> 57

gagagctagc atgccggcta gatctcgaga tcggccggcc tagg

44

<210> 58

<211> 30

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence; note =  
synthetic construct

<400> 58

gaactgcagc tcgagagcta gcatgccggc

30

<210> 59

<211> 32

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence; note =  
synthetic construct

<400> 59

ggagatatac atatggctag catgactggt gg

32

<210> 60

<211> 31

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence; note =  
synthetic construct

<400> 60

tgctcgaagc ttcgcagaag ataatagtag g

31